A WIDER VIEW OF THE RELATIONSHIP BETWEEN SETTLEMENT SIZE AND POPULATION IN THE PERUVIAN ANDES

Kenneth L. Khamme

Although the relationship between settled area and population size has been well studied in archaeology, anthropology, and geography, with numerous empirical data sets, an established methodological approach, and a body of theory. Schreiber and Kintigh’s (1996) analysis of archaeological and historical data from the Peruvian Andes found “only a weak correlation” between these variables. By employing double logarithmic data, the relationship is shown to be much better than the “surprisingly poor” one the raw data suggests. Regression analysis of these Peruvian data in logarithmic form yields power functions that conform closely with expectations derived from a body of settlement size theory based on the allometric principle.

La relación entre área habitacional y tamaño poblacional ha sido bien estudiada en arqueología, antropología, y geografía, con numerosas muestras empíricas, un acercamiento metodológico establecido, y amplia literatura teórica. Sin embargo Schreiber y Kintigh (1996) han encontrado en su análisis de datos arqueológicos e históricos de los Andes peruanos “sólo una débil correlación” entre estas variables. Empleando datos logarítmicos dobles la relación resulta mucho mayor que la “sorprendentemente menor” sugerida por los datos. El análisis de regresión de estos datos peruanos en forma logarítmica proporciona funciones de potencia que se conforman muy bien con las predicciones derivadas de la literatura de teoría de patrones de asentamiento basada en el precepto alométrico.

Schreiber and Kintigh (1996) recently investigated the relationship between settlement area and population size in the central highlands of Peru using census data from 1540. Their paper provides an excellent example of a problem context that is particularly well suited for quantitative treatment, and one that has an unusually rich history within and outside of archaeology. After all, if archaeologists can develop reliable means to estimate population sizes based on the areas of past settlements, a potentially powerful tool would be realized. Schreiber and Kintigh’s presentation also illustrates exemplary caution and thoughtfulness in approaching the analysis of the data—by focusing first on scatter plots and considering the effects of outliers, for example.

Yet, despite their prudence, Schreiber and Kintigh overlooked a fundamental Exploratory Data Analysis (EDA) tactic, a rich theoretical literature, and empirical findings that have been made in this area by anthropologists, archaeologists, and geographers alike (e.g., Cook and Heizer 1968; Naroll 1962; Nordbeck 1971; Ogrosky 1975; Stewart and Warntz 1968; Tobler 1969; Wiessner 1974). Surprisingly, what is called for by EDA, theory, and previous empirical case studies happens to be closely linked.

A key “problem” in Schreiber and Kintigh’s data analysis revolves around what to do with the outlier of Apcara, the largest settlement with more than twice the area and population of any of the other sites considered. Following one data analysis principle, Schreiber and Kintigh consider the strong effect this single case would have in correlation/regression analyses and eliminate it in some. They do not consider another fundamental data analysis tactic that is central to the EDA school, however: re-expression of the data in logarithmic form (e.g., see Mosteller and Tukey 1977:Chapter 5). This is a common approach utilized by data analysts. Thomas (1986:434) notes that our reluctance to transform, or our oversight in doing so, reflects mere habit or convention in the use of raw measurement scales.

Logarithmically transformed data tend to be “better behaved”; they show less skewness, and the extremity of outlying data points is reduced. This is clearly illustrated in Figure 1 where the...
large outlying settlement (Apaca) in the raw data is brought in much closer to the pack, and the data in general show a much less skewed distribution when plotted in double logarithmic form. The literature (e.g., Cook and Heizer 1968; Naroll 1962; Nordbeck 1971; Ogrosky 1975) shows this data pattern to be the common case where a relatively small number of large settlements exist on the right tail of the size distribution; the log transform therefore has generally proven essential in analyses of these kinds of data. (The same may also be said of studies that examine the rank-size relationship between settlements [e.g., see Hodder and Orton 1976:69].)

The log transform gives other advantages. In a bivariate regression a linear function is merely a special case of the logarithmic. If a log-log plot shows a slope of 1.0, then the raw plot will be linear. But if the slope is any other value, the raw plot will produce a curve. The log form therefore subsumes the linear expression. Finally, in log form the slope becomes a dimensionless index (not so in the raw case) that describes the relationship between x and y as a functional rate of change. In examining population size as a function of settlement area, a slope of less than unity indicates increasingly more area per person as settlement size gets larger. In other words, population may not be a constant fraction of area as one assumes in a linear model. This is exactly what Schreiber and Kintigh’s data show when examined collectively (Table 1, where the slope of a model for all settlements is $b = .5004$), and seems to be the rule in studies of this sort (cited previously).

Theoreticians (e.g., Naroll 1962; Nordbeck 1971; Wiessner 1974) have therefore suggested that the settlement area-population size relationship is merely another reflection of the allometric growth principle seen in biological organisms, where it is the rate of relative change in a variable y (e.g., the growth rate of an organ) that is a constant fraction of the rate of relative change in another variable x (e.g., an organism’s size). Such relationships are linear only in logarithmic form (see Thomas 1986:434–436 for an overview of allometric relationships).

In logarithmic form the regression model, $\log(y) = \log(a) + b\log(x)$, is usually reexpressed as the power function: $y = ax^b$ (where y = population and $x = area$). Considering Schreiber and Kintigh’s 11 settlements as a single group yields the model parameters a and b listed in Table 1 ($R^2 = .66, r = .81, p = .003$). With some justification Schreiber and Kintigh go on to partition their data set into two groups: one of four regional centers and the other of seven villages. In doing so, they achieved much stronger linear fits with the data in each of the groups (although they were still faced with some skewness in the distributions and the outlier of Apaca). These models yielded coefficients of
determination of $R^2 = .81$ ($p = .006$) for the villages and $R^2 = .99$ ($p = .001$) for the centers. Models based on the "better behaved" double-logarithmic data (Figure 1) yield respective $R^2$ statistics of .76 and .99 for the same groups ($p \leq .01$ in both cases), with model coefficients listed in Table 1.

Most settlement size theory has focused on area as a deterministic function of population size (I hazard to present this with the fall-from-grace of such perspectives in much contemporary archaeology). Schreiber and Kintigh’s data, perhaps coincidentally, neatly complement this theory, and I present these findings as a curiosity. Nordbeck (1971) initially pointed out that in contemporary urban contexts the settled area is of dimension two, but the population distribution is of dimension three (analogous to the shape of a volcano with a ring of high population—presumably vertical apartment dwellings—around a low-density urban core). He then deduced that the exponent, $b$, in the allometric formula: $area = (a)population^b$, must therefore be $\frac{2}{3}$, and went on to show this to be the case using Swedish data.

Wiessner (1974) attempted to generalize this theory to other settlement types. Observing that hunter-gatherer population distributions are distributed in a line of huts that form the perimeter of a camp circle (using the !Kung Bushman as a model), she reasoned that the exponent, $b$, in these contexts should be $\frac{2}{3}$ (the dimension of the dependent variable, area, is always the numerator; the dimension of the independent variable, population, is the denominator). A regression using available !Kung data indicated close agreement. Viewing the village settlement type on a continuum between the hunter-gatherer camp and the urban center, Wiessner (1974:349) suggested that villages “with more or less . . . evenly filled areas, but little vertical dimension” should yield an areal distribution with an exponent of $\frac{2}{3} = 1.0$, but did not go on to test this relationship.

Using Schreiber and Kintigh’s logarithmically transformed data, this time with settlement area as the dependent variable, gives an exponent (b) of .99 for the centers ($R^2 = .99, p = .002$) and .84 for the villages ($R^2 = .76, p = .01$; Table 1). If we remove the two villages with boundaries that may have constrained “normal” growth (Paracha and Sondondo), as Schreiber and Kintigh (1996:578) did, the exponent becomes 1.02 ($R^2 = .97, p = .002$) for the remaining five villages of this class (I leave it to the reader to attribute any higher meaning to these results).

Finally, it should be noted that the intercept or $a$-values of the previous equations are also informative because they form what Nordbeck (1971:58) terms a “space standard.” For the four centers $a = .1479$ and for the five unconstrained villages $a = .0431$ (Table 1). This indicates about .1479 ha/person (1.479 m$^2$) for the centers and .0431 ha/person (431 m$^2$) for villages, conforming with Schreiber and Kintigh’s (1996:577) perceptions of the use of space in this cultural context.

Acknowledgments. The suggestions of three anonymous reviewers and the editor substantially improved the quality of this paper.

References Cited


Naroll, R.

Nordbeck, S.

Ogrosky, C. E.

Schreiber, K. J., and K. W. Kintigh

Stewart, J. Q., and W. Warmt

Thomas, D. H.

Tobler, W. R.

Wiessner, P.

Received December 6, 1996; accepted February 28, 1997; revised April 7, 1997.

---

**Economy and Society in Prehistoric Europe**

**Changing Perspectives**

**Andrew Sherratt**

Over the past twenty-five years Andrew Sherratt has reinterpreted the growing mass of information about the economic foundations of prehistoric Europe using concepts drawn from history and the social sciences. Sherratt’s classic articles on the subject are collected together here for the first time. Sherratt identifies in his commentaries the significance of each article within the framework of the entire collection. In addition to these commentaries, the collection includes notes throughout the text that update the reader on developments taking place in the field after the time of each article’s first publication.

Cloth $59.50 ISBN 0-691-01697-6

---

"Andrew Sherratt is among the foremost authorities in European prehistory and this important volume of essays covers the whole range of his thinking. . . . It is a work of significance for all students of the early past of Europe."

—Colin Renfrew

---

**Princeton University Press**

AVAILABLE AT FINE BOOKSTORES OR DIRECTLY FROM THE PUBLISHER: 800-777-4726

VISIT OUR WEBSITE: PUP.PRINCETON.EDU