EMPIRICAL ANALYSIS OF SEGREGATION INDEXES (1)

By: Karl E. Taeuber, National Cancer Institute

Residential segregation of whites and nonwhites exists in all U. S. cities. Obviously no one measure can represent, adequately and for all purposes, the patterns of residential segregation in a city. Several "segregation indexes" have been proposed to serve as general summary measures (2). This paper reports on some empirical comparisons of these indexes.

A distinction may be maintained between the behavioral processes involved in residential segregation, and the resulting spatial distribution of white and nonwhite residents. The various segregation indexes share a general approach in the specification of a measurable aspect of the spatial distribution. Several aspects of the spatial patterns might be of interest. These include the extent to which white or nonwhite residents are clustered together, the location within the city or metropolitan area of such clusters, and the general evenness in the distribution of white and nonwhite residents throughout the city. The segregation indexes share a common focus on this latter aspect.

For purposes of measurement, the concept of general evenness in the spatial distribution may be clarified by defining the situations of complete evenness and complete unevenness. Consider a city divided into subareas, such as tracts, wards, or blocks. If each subarea is occupied entirely either by whites or by nonwhites, and no subarea contains residents of both colors, the spatial distribution is said to be completely uneven. A value of 1 may be assigned to this situation, representing maximum residential segregation. If in each subarea nonwhite residents constitute the same proportion of all residents as they do in the entire city, the distribution is said to be completely even. A value of 0 may be assigned to this situation, representing minimum residential segregation.

All of the proposed segregation indexes derive from this general approach. The basic data from which any of the indexes may be computed are the numbers of white and nonwhite residents in each subarea. There are two ways in which indexes computed for a given city may yield different values. The subareas utilized for computation may differ, or the formulas for combining the distributional data into a summary figure may differ. The indexes are all measures of the same general concept, are computed from the same type of data, and share common maximum and minimum values. How important are the differences between index values that may arise from the two sources noted?

It is readily apparent that the choice of subarea will affect the value of a particular index. In general, the smaller the subarea, the greater the unevenness that will be noted, and the higher the index value. For instance, nonwhite residents might occupy approximately ten percent of the dwellings in each census tract, and yet within each tract be clustered together in block occupied only by nonwhites. Such a situation could result in a tract index close to 0 and a block index of 1.

Census tracts and blocks are the two subareas for which data are most readily available. The use of census tracts for computing segregation indexes has been criticized (3), but the general arguments may apply to any areal unit. The boundaries of an areal unit may or may not be useful in delineating residential areas so as to indicate the full extent of residential separation. In delineating census tracts, some tract committees have tried to make tract boundaries correspond with boundaries between white and nonwhite residential areas. Other tract committees have paid relatively little attention to race. In either case, over time the degree of correspondence between tract boundaries and boundaries of white and nonwhite settlement may shift. Clusters of nonwhite occupied dwellings may overlap tract boundaries, or be entirely included within individual tracts. Hence it has been argued that the large size and the arbitrary delineation of tract boundaries mitigate against their use for the computation of segregation indexes. Except that blocks are not generally delineated by committees, the same arguments apply. If nonwhites tend to live in alleys or on side streets, while whites live along the main street fronts, then sole reliance on block data may also be misleading.

Hypothetical argument cannot indicate the empirical importance of the areal factor. It is plausible to argue that extensive concern with the relative merits of blocks and tracts may not be necessary. Although the specific magnitudes of tract and block indexes differ, how much effect does the choice of areal unit have on the comparative rankings of a number of cities?

Comparisons were made for each of 6 different segregation indexes. For five of the indexes, comparisons are between values computed from tract data and those computed from block data for 60 cities, using 1940 data. For the sixth index, the comparisons are based on data for 72 cities for 1950 (4). Table 1 gives the product-moment correlation coefficients between the block and tract values for each of the six indexes. Because tracts are fewer in number, tract indexes are more easily computed. Hence it might be desired to use tract indexes for estimating the values of block indexes. Table 1 also includes the parameters of the regressions of block on tract indexes.

For five of the indexes, there is a definite linear relationship between the block and tract values. The sixth index, Co, is the proportion of all whites in the city who reside in subareas that are exclusively white-occupied. If segregation is defined with reference only to this one aspect of the spatial distribution,

the index values and interpretations from them will differ markedly according to the type of subarea used in the computations.

The three indexes, Gh, Rep, and Bell, have in common a distinct dependence on the city nonwhite proportion. The proportion of nonwhites in a city remains the same, whether blocks or tracts are used for computing the indexes. This dependence on a common factor is a partial explanation of the high correlations between block and tract values of these indexes.

The indexes, Gi and D, are, conceptually, the most adequate measures of overall evenness, without specific reference to the proportion of nonwhites. The correlations of block and tract values for these indexes, .74 and .69, indicate only a moderate relationship.

The deviations of particular cities from the regressions of block on tract indexes vary with the index. The differences between block and tract values are functions of the index as well as of the spatial distribution. For cities with large numbers of tracts, the block indexes are only slightly more highly correlated with tract indexes than is the case for cities with few tracts.

For those indexes with a strong dependence on the proportion of nonwhites in the total population (Gh, Rep, Bell), the difference between block and tract values are relatively slight, and intercity comparisons based on either subarea will yield similar results. For indexes that are pure measures of unevenness (Gi, D, Co), the values computed from tract data are an unreliable guide to those computed from block data.

The conclusions derived from studying the rank correlations of tract and block indexes are essentially the same. For gross intercity comparisons of the level of segregation on a given index, block or tract values may be used without greatly affecting the comparisons (except for Co). For more precise analysis, the conclusions will depend on whether tracts or blocks are chosen as subareas for index computation.

The segregation indexes differ in the items of information they utilize and in the formulas by which values between 0 and 1 are assigned to various residential distributions. The Duncans have reported on some of the index properties and on certain algebraic interrelations among the indexes. Thus, some of the indexes utilize the proportion of nonwhites in the city population, while others are pure measures of unevenness in that they utilize only the data on the spatial distribution. One index, the original Cowgill index (Cwg), is distinctive in the definition of maximum and minimum segregation. The extreme distributions are rarely encountered, however, and such a conceptual difference might not result in an empirical difference between the Cowgill index and other indexes. The general question may be raised whether the various differences between the concepts and definitions of the indexes are of significance in empirical investigation. Will a city be found to have approximately the same degree of residential segregation, as compared to other cities, regardless of which segregation index is used?

Table 2 gives the intercorrelations of 8 segregation indexes computed from 1950 block data for 188 cities. The numerous low coefficients indicate that the indexes are not simply minor variants of each other. Rather, the algebraic differences noted by the Duncans tend to be evident in the pattern of intercorrelations.

The first column of the table shows the correlations between each index and the proportion of nonwhites in the city population. The coefficients are positive and moderately high for the three indexes (Gh, Rep, Bell) directly utilizing the proportion of nonwhites in their formulas. Two pure measures of spatial unevenness (Gi, D) have very little relationship to the proportion of nonwhites. This difference between the indexes is obviously of major importance in any substantive analysis of residential segregation.

The index which deviates from the others in the definition of maximum and minimum segregation (Cwg) is not highly correlated with any of the other indexes. The three indexes, Gh, Rep, and Bell, are highly intercorrelated with each other. The nearly perfect linear relationship between Gh and Bell confirms the impression that the Bell index is of the same general type as the other two.

The indexes Co and Oc, although pure measures of spatial unevenness, each depend on one aspect of the distribution. They are empirically distinctive from each other and from each of the other indexes. The other pure measures, Gi and D, behave quite similarly. If only those cities with large Negro populations are considered, the similarities between Gi and D are greater.

Although the indexes are not empirically identical, they are not 8 completely independent measures. Using tract data, the Duncans have reported that observed values of D and the proportion of nonwhites could be used to make quite good estimates of Gi, Gh, Rep, and Bell. It is clear from Table 2 that good predictions could be made of some of the block indexes, given values for the others. The proportion of nonwhites and indexes D and Cwg are relatively independent of each other. Without precise evaluation, it is clear that linear combinations of these three could be used to estimate values of the other six indexes. At least 60 percent of the variance in Oc could be accounted for, and better predictions would result for the others.

Another means of analyzing the interdependence of the nine measures is factor analysis. A brief exploration was carried out,

using the Turinsky direct factor program for Univac. The raw data were used, and the resulting factors are not readily identifiable. However, the patterning of the factor loadings on the indexes for the four extracted factors confirms the previous analyses. Gi and D have similar loadings. Cwg stands by itself, except for similar loadings to Co on two factors. Co is close to Gi and D, the other pure measures, on two factors. Oc, however, tends to be similar to the proportion of nonwhites. Bell, Gh, and Rep form a group with similar factor loadings.

The preceding analyses of the empirical behavior of segregation indexes demonstrate very important differences between them. Values of the indexes computed for U. S. cities may differ because the areal units from which they are computed differ, or because of differences in the ways the areal data are utilized. The results of any substantive analysis utilizing segregation indexes will therefore depend both on the index chosen and on the areal unit data from which it is computed.

For some empirical investigations of residential segregation, tract data may have particular advantages over block data. For inter-city analysis of the spatial evenness in the residential distribution, however, block data are generally preferable. Blocks are smaller and less subject to arbitrary delineation. These properties increase the comparability of indexes computed for different cities. Although perfect comparability is not possible with indexes based on areal data, blocks are of sufficiently fine scale that the limitations to comparability are not serious. In addition, block data are more sensitive than tract data to small variations in the evenness of the residential distribution. Block boundaries are more stable over time. Block data are available for both 1940 and 1950 for more cities, although tract data are available for some entire metropolitan areas.

The choice of a particular index must be based on the purposes of the investigation. For most analyses, it is desirable to study separately the spatial evenness and the proportion of nonwhites. The compounding of these two dimensions in a single index is thus an undesirable property. Of the pure measures of evenness, it has already been noted that Gi and D are the most adequate central measures of overall evenness. Although the Gini index is more sensitive to all parts of the spatial distribution, the dissimilarity index (D) is empirically quite similar and is more easily computed. For intercity comparisons of spatial unevenness in the distribution of white and nonwhite residents, either the Gini index or the dissimilarity index, computed from block data, is likely to be the analytically most useful choice (5).

Table 1. Block and Tract Indexes of Residential Segregation between Whites and Nonwhites: Means, Correlations, and Regressions.

Measure	Index and Year					
	Gi 1940	D 1940	Co 1940	Gh 1940	Rep 1940	Bell 1950
Number of Cities	60	60	60	60	60	72
Mean of Block Indexes	.96	.86	.74	.70	.51	.64
Mean of Tract Indexes	.82	.70	.08	.50	.29	.44
Correlation of Block and Tract Indexes	.74	.69	.63	.92	.90	.93
Regression of Block on Tract Indexes						
Slope	.18	.31	.78	.70	.93	.89
Intercept	.81	.64	.68	.35	.24	.25

Table 2. Intercorrelations of Eight Block Indexes of Residential Segregation between Whites and Nonwhites and the Proportion of Nonwhites, 188 Cities, 1950.

	q	Gi	D	Co	Oc	Gh	Rep	Bell
Gi	.19							
D	-.01	.90						
Co	-.50	.58	.75					
Oc	.76	.51	.32	-.13				
Gh	.69	.49	.13	-.32	.75			
Rep	.79	.52	.22	-.31	.85	.95		
Bell	.72	.52	.19	-.30	.81	.99	.98	
Cwg	.20	.55	.05	-.04	.34	.70	.49	.63

For index definitions and references, see Footnote (2). The letter q represents the proportion of nonwhites to total city population.

Footnotes:

(1) This is Paper No. 5 in the series, "Comparative Urban Research", issuing from the Population Research and Training Center, University of Chicago, under a grant from the Ford Foundation. Computation of the indexes was supported by the Laboratory of Social Relations, Harvard University. The research was performed under a National Science Foundation Predoctoral Fellowship.

(2) For definitions of the indexes, analysis of their algebraic interrelations, additional discussion of some of the problems considered here, and references to other sources, see: Otis Dudley Duncan and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," American Sociological Review, 20(1955), 210-217. Cwg, the Cowgill index, is defined in the reference in footnote (3). Oc is a variation of the other indexes, and is equal to the proportion of all nonwhites in the city who reside in subareas occupied exclusively by nonwhites.

(3) Donald O. Cowgill and Mary S. Cowgill, "An Index of Segregation Based on Block Statistics," American Sociological Review, 16(1951), 825-831.

(4) The five indexes based on 1940 tract data were computed by the Duncans. Community Areas were used for Chicago and Health Areas for New York. The Bell indexes for 1950 tract data are given in: Wendell Bell and Ernest M. Willis, "The Segregation of Negroes in American Cities," Social and Economic Studies, 6(1957), 59-75. All block indexes except the Cowgill index were computed by the author; see the reference in (5).

(5) The research reported on here is discussed in greater detail in the author's unpublished doctoral dissertation, "Residential Segregation by Color in United States Cities, 1940 and 1950," submitted to Harvard University, December, 1959.