

# MEASURES OF RESIDENTIAL SEGREGATION

Massey and Denton (1988) used an extensive literature search and cluster analysis to identify 20 different indexes of segregation and classify them into five key dimensions of segregation. Basically, **evenness** involves the differential distribution of the subject population, **exposure** measures potential contact, **concentration** refers to the relative amount of physical space occupied, **centralization** indicates the degree to which a group is located near the center of an urban area, and **clustering** measures the degree to which minority group members live disproportionately in contiguous areas. Below we describe the 19 measures we have calculated, though we focus on only one per dimension, as described in Chapter 2. In all of the calculations, non-Hispanic Whites are considered the “majority” (reference) population. The formulas for the 19 indexes are listed at the end of this appendix.

## I. MEASURES OF EVENNESS

Evenness measures of segregation compare the spatial distributions of different groups among units in a metropolitan area. Segregation is

smallest when majority and minority populations are evenly distributed. The most widely used measure of evenness is the **dissimilarity** index. Conceptually, dissimilarity measures the percentage of a group’s population that would have to change residence for each neighborhood to have the same percentage of that group as the metropolitan area overall. The index ranges from 0.0 (complete integration) to 1.0 (complete segregation).

A second measure of evenness is the **Gini** coefficient. Like the index of dissimilarity, it can be derived from the Lorenz curve, and varies between 0.0 and 1.0, with 1.0 indicating maximum segregation. The Gini coefficient is “the mean absolute difference between minority proportions weighted across all pairs of areal units, expressed as a proportion of the maximum weighted mean difference” (Massey and Denton, p. 285). A third evenness measure is **entropy**, proposed originally by Theil (Theil 1972; Theil and Finizza, 1971). The entropy index (also called the information index) measures the (weighted) average

deviation of each areal unit from the metropolitan area’s “entropy” or racial and ethnic diversity, which is greatest when each group is equally represented in the metropolitan area. The entropy index, like the other two evenness measures, also varies between 0.0 (when all areas have the same composition as the entire metropolitan area) and 1.0 (when all areas contain one group only).

The only evenness measures to satisfy the four criteria established by James and Taeuber (1985) for an ideal segregation index<sup>1</sup> are the Gini index and the **Atkinson** index, often used to evaluate inequality.<sup>2</sup> The Atkinson index

<sup>1</sup> The four criteria are: 1) the “transfer principle”, which states that a measure should be sensitive to the redistribution or “transfer” of minorities among areal units with minority proportions above or below the metropolitan area’s minority proportion (and not just transfers from areas above to areas below that proportion); 2) “compositional invariance”, which states that the relative size of minority population should not affect the index; 3) “size invariance”, which states that the measure should not be affected if the number of people in each group is multiplied by a constant; and 4) “organizational equivalence”, which holds that an index should be unaffected by aggregating units with the same minority composition.

<sup>2</sup> See Jones and Weinberg (2000) for an application to income inequality.

(Atkinson, 1970) allows the researcher to differentially weight areal units at different points along the Lorenz curve, allowing, for example, areal units where minorities are under- or over-represented to contribute more heavily to the overall index. For values of the shape parameter of 0.0 or more but less than 0.5, areal units where the proportion of minorities is smaller than the metropolitan area's average (i.e., where minorities are "underrepresented") contribute more to the segregation index; for large values of the shape parameter (more than 0.5 up to 1.0), the reverse is true — areas of "overrepresentation" contribute more. When the shape parameter is 0.5, such areas contribute equally. For values of the parameter between 0.0 and 1.0, the Atkinson index also varies in that range, with 1.0 indicating maximum segregation. Three shape parameters — 0.1, 0.5, and 0.9 — are used in our analysis, resulting in three separate Atkinson indexes.

## II. MEASURES OF EXPOSURE

"Exposure measures the degree of potential contact, or possibility of interaction, between minority and majority group members" (Massey and Denton, p. 287). Exposure thus depends on the extent to which two groups share common residential areas, and hence, on the degree to which the average minority group member "experiences" segregation. As Massey and Denton point out, indexes of evenness and exposure are correlated but measure different things: exposure measures depend on the relative sizes of the two groups being compared, while evenness measures do not.

The two basic, and related, measures of exposure are **interaction** and **isolation**. The two indexes, respectively, reflect the probabilities that a minority person shares a unit area with a majority person or with another minority person. The interaction index measures the exposure of minority group members to members of the majority group as the minority-weighted average of the majority proportion of the population in each areal unit. The isolation index measures "the extent to which minority members are exposed only to one another," (Massey and Denton, p. 288) and is computed as the minority-weighted average of the minority proportion in each area.

When there are only two groups, the isolation and interaction indexes sum to 1.0, so lower values of interaction and higher values of isolation each indicate higher segregation. However, when there are more than two groups, the interaction and isolation indexes will not sum to 1.0 (one must add the interaction indexes for all other minority groups to the interaction and isolation indexes for the original minority group to obtain unity).<sup>3</sup> Furthermore, the interaction indexes representing minority exposure to majority members and majority exposure to minority members will be equal only if the two groups constitute the same proportion of the population. An adjustment of the isolation index to control for this asymmetry yields a third exposure index, the **correlation** ratio, also known as eta-squared.

<sup>3</sup> The interaction and isolation indices reported here are calculated with the non-Hispanic White population as the referent group rather than the total population (excluding the minority group of interest). Our interaction and isolation indices therefore do sum to unity for each group.

## III. MEASURES OF CONCENTRATION

"Concentration refers to the relative amount of physical space occupied by a minority group in the metropolitan area" (Massey and Denton, p. 289). Minority groups of the same relative size occupying less space would be considered more concentrated and consequently more segregated.

One measure of concentration, originally proposed by Hoover (1941), is **delta**, which "computes the proportion of [minority] members residing in areal units with above average density of [minority] members" (Massey and Denton, p. 290). The index gives the proportion of a group's population that would have to move across areal units to achieve a uniform density.

Massey and Denton propose two additional measures. The first, **absolute concentration**, computes the total area inhabited by a group and compares this with the minimum and maximum areas (the areal sum, respectively, of the fewest number of the geographically smallest and the greatest number of the geographically largest areal units) that could accommodate a group of that size (at observed densities). The index varies from 0.0 to 1.0, where a score of 1.0 means that a group has achieved the maximum spatial concentration possible (all minority members live in the smallest areal units). The second, **relative concentration**, is measured similarly, but takes account of the distribution of the majority group as well. This measure varies from -1.0 to 1.0.<sup>4</sup> A score of 0.0 means that the

<sup>4</sup> In fact we obtained values below -1.0, and Massey and Denton (personal communication) indicate that they did as well in calculating indices for the 60 largest metropolitan areas.

minority and majority groups are equally concentrated. An index of -1.0 means that the concentration of the majority exceeds that of the minority to the maximum extent, and an index of 1.0 the reverse.

#### IV. MEASURES OF CENTRALIZATION

“Centralization is the degree to which a group is spatially located near the center of an urban area” (Massey and Denton, p. 291). As for concentration, absolute and relative measures are presented.

**Relative centralization** compares the areal profile of the majority and minority populations, and may be interpreted as the relative share of the minority population that would have to change their area of residence to match the centralization of the majority. The index varies between -1.0 and 1.0 with positive values indicating that minority members are located closer to the center than majority, and negative values the reverse. An index of 0.0 indicates that the two groups have the same spatial distribution around the center.

**Absolute centralization** examines only the distribution of the minority group around the center and also varies between -1.0 and 1.0. “Positive values indicate a tendency for [minority] group members to reside close to the city center, while negative values indicate a tendency to live in outlying areas. A score of 0 means that a group has a uniform distribution throughout the metropolitan area” (Massey and Denton, p. 293).

Most analysts using a centralization measure define it in terms of access to the traditional Central Business District (CBD). We feel that this concept is increasingly outmoded as jobs, retail sales, and

other CBD functions continue to decentralize. Similarly, the transitional zones that once developed around CBDs should, following Hoover’s (1941) logic, also develop around many of the “multiple nuclei” that also have partially supplanted CBDs (see Garreau, 1991). Accordingly, we have defined an alternative to CBD-centered indexes that are based on distance from the population centroid.

#### V. MEASURES OF CLUSTERING

Clustering measures “the extent to which areal units inhabited by minority members adjoin one another, or cluster, in space” (Massey and Denton, p. 293). A high degree of clustering indicates a racial or ethnic enclave. The first measure of clustering is **absolute clustering**. This index “expresses the average number of [minority] members in nearby [areal units] as a proportion of the total population in those nearby [areal units]”, where distances between areal units are measured from their centroids (Massey and Denton, p. 294). It varies from 0.0 to 1.0.<sup>5</sup>

White’s (1986) index of **spatial proximity** is the average of intra-group proximities for the minority and majority populations, weighted by the proportions each group represents of the total population. Spatial proximity equals 1.0 if there is no differential clustering between minority and majority group members. It is greater than 1.0 when members of each group live nearer to one another than to members of the other group, and is less than 1.0 if minority and

majority members live nearer to members of the other group than to members of their own group.

Massey and Denton derive from these two measures an index of **relative clustering**, which “compares the average distance between [minority] members...with the average distance between [majority] members” (Massey and Denton, p. 295). The index equals 0.0 when minority members display the same amount of clustering as the majority, is positive when minorities display greater clustering than the majority, and is negative if they are less clustered than the majority.

If there is clustering, the number of majority persons with whom a minority might potentially interact should increase with increasing distance from the minority’s area of residence. However, the likelihood of actual encounters and interaction with majority persons should decay rapidly. The **distance-decay interaction** index measures this as the sum of the probabilities that a minority person in each tract *i*, weighted by the minority proportion in that tract, would encounter a resident in another tract *j*, weighted by the proportion of majority persons in tract *j*. The index can be interpreted as measuring the probability that the next person a minority group member meets anywhere in the city is a majority member.

Corresponding to this is a **distance-decay isolation** index, which measures the probability that the person a minority next encounters is also a minority. Massey and Denton note that Morgan’s (1983) paper proposing these distance-decay indexes did not describe their behavior or provide an empirical example. However, as a distance

<sup>5</sup>We have obtained negative values, though these have been close to zero. Rounding error in calculations could play a role.

weighted function of the exposure interaction and isolation indexes, one might expect the distance-decay measures to also to vary between 0.0 and 1.0, with 0.0 representing maximum segregation on the distance-decay interaction index and 1.0 indicating this on the distance-decay isolation index. The values obtained from census data suggest such a range.

## VI. TECHNICAL DESCRIPTION OF SEGREGATION INDEXES

### Definitions

- n the number of areas (census tracts) in the metropolitan area, ranked smallest to largest by land area
- m the number of areas (census tracts) in the metropolitan area, ranked by increasing distance from the Central Business District ( $m = n$ )
- $x_i$  the minority population of area  $i$
- $y_i$  the majority population (non-Hispanic Whites in this report) of area  $i$
- $y_j$  the majority population of area  $j$
- $t_i$  the total population of area  $i$
- $t_j$  the total population of area  $j$
- $X$  the sum of all  $x_i$  (the total minority population)
- $Y$  the sum of all  $y_i$  (the total majority population)
- $T$  the sum of all  $t_i$  (the total population)
- $p_i$  the ratio of  $x_i$  to  $t_i$  (proportion of area  $i$ 's population that is minority)
- $P$  the ratio of  $X$  to  $T$  (proportion of the metropolitan area's population that is minority)
- $a_i$  the land area of area  $i$
- $A$  the sum of all  $a_i$  (the total land area)
- $n_1$  rank of area where the sum of all  $t_i$  from area 1 (smallest in size) up to area  $n_1$  is equal to  $X$
- $T_1$  the sum of all  $t_i$  in area 1 up to area  $n_1$
- $n_2$  rank of area where the sum of all  $t_i$  from area  $n$  (largest in size) down to area  $n_2$  is equal to  $X$
- $T_2$  the sum of all  $t_i$  in area  $n_2$  up to area  $n$
- $d_{ij}$  the distance between area  $i$  and area  $j$  centroids, where  $d_{ij} = (0.6a_i)^{0.5}$
- $c_{ij}$  the exponential transform of  $-d_{ij}$  [ $= \exp(-d_{ij})$ ]
- $b$  a shape parameter that determines how to weight the increments to segregation contributed by different portions of the Lorenz curve

### Index Formulas

Note: Indexes in this report were calculated as if non-Hispanic Whites and the minority group in question were the only two groups present in the total population. Formulas are from Massey and Denton (1988).

### 1. Dissimilarity

$$\frac{\sum_{i=1}^n \left[ t_i \left| (p_i - P) \right| \right]}{[2TP(1 - P)]}$$

### 2. Gini

$$\frac{\sum_{i=1}^n \sum_{j=1}^n t_i t_j \left| (p_i - p_j) \right|}{2T^2 P(1 - P)}$$

### 3. Entropy

$$\sum_{i=1}^n \left[ \frac{t_i (E - E_i)}{ET} \right]$$

where  $E_i = p_i \ln \left( \frac{1}{p_i} \right) + (1 - p_i) \ln \left( \frac{1}{1 - p_i} \right)$

and  $E = P \ln \left( \frac{1}{P} \right) + (1 - P) \ln \left( \frac{1}{1 - P} \right)$

### 4. Atkinson (parameter $b$ )

$$1 - \left( \frac{P}{1 - P} \right) \left[ \sum_{i=1}^n \left[ \frac{(1 - p_i)^{1-b} p_i^b t_i}{PT} \right] \right]^{\frac{1}{1-b}}$$

### 5. Interaction

$$\sum_{i=1}^n \left[ \left( \frac{x_i}{X} \right) \left( \frac{y_i}{t_i} \right) \right]$$

### 6. Isolation

$$\sum_{i=1}^n \left[ \left( \frac{x_i}{X} \right) \left( \frac{x_i}{t_i} \right) \right]$$

### 7. Correlation

$$\frac{(I - P)}{(1 - P)}$$

where  $I$  is the isolation index

8. Delta

$$0.5 \sum_{i=1}^n \left| \left( \frac{x_i}{X} \right) - \left( \frac{a_i}{A} \right) \right|$$

9. Absolute Concentration

$$1 - \frac{\left[ \sum_{i=1}^n \left( \frac{x_i a_i}{X} \right) - \sum_{i=1}^{n1} \left( \frac{t_i a_i}{T_1} \right) \right]}{\left[ \sum_{i=n2}^n \left( \frac{t_i a_i}{T_2} \right) - \sum_{i=1}^{n1} \left( \frac{t_i a_i}{T_1} \right) \right]}$$

10. Relative Concentration

$$\left\{ \frac{\left[ \sum_{i=1}^n \left( \frac{x_i a_i}{X} \right) \right] - 1}{\left[ \sum_{i=1}^n \left( \frac{y_i a_i}{Y} \right) \right] - 1} \right\} \left\{ \frac{\left[ \sum_{i=1}^{n1} \left( \frac{t_i a_i}{T_1} \right) \right] - 1}{\left[ \sum_{i=n2}^n \left( \frac{t_i a_i}{T_2} \right) \right] - 1} \right\}$$

11. Absolute Centralization

$$\sum_{i=1}^m (X_{i-1} A_i) - \sum_{i=1}^m (X_i A_{i-1})$$

12. Relative Centralization

$$\sum_{i=1}^m (X_{i-1} Y_i) - \sum_{i=1}^m (X_i Y_{i-1})$$

13. Absolute Clustering

$$\left\{ \sum_{i=1}^n \left[ \frac{x_i}{X} \sum_{j=1}^n c_{ij} x_j \right] - \left[ \frac{X}{n^2} \sum_{i=1}^n \sum_{j=1}^n c_{ij} \right] \right\} \left\{ \sum_{i=1}^n \left[ \frac{x_i}{X} \sum_{j=1}^n c_{ij} t_j \right] - \left[ \frac{X}{n^2} \sum_{i=1}^n \sum_{j=1}^n c_{ij} \right] \right\}$$

14. Spatial Proximity

$$\frac{(X P_{xx} + Y P_{yy})}{T P_{tt}}$$

where  $P_{gg} = \sum_{i=1}^n \sum_{j=1}^n \left[ \frac{(g_i g_j c_{ij})}{G^2} \right]$

and  $\{g, G\} = \{x, X\}, \{y, Y\}, \{t, T\}$

15. Relative Clustering

$$\left( \frac{P_{xx}}{P_{yy}} \right) - 1$$

16. Distance-Decay Interaction

$$\sum_{i=1}^n \left[ \frac{x_i}{X} \sum_{j=1}^n \left( \frac{k_{ij} y_j}{t_j} \right) \right]$$

where  $k_{ij} = \left[ \frac{t_j^{(-d_{ij})}}{\sum_{i=1}^n t_{ij}^{(-d_{ij})}} \right]$

17. Distance-Decay Isolation

$$\sum_{i=1}^n \left[ \frac{x_i}{X} \sum_{j=1}^n \left( \frac{k_{ij} x_j}{t_j} \right) \right]$$