

## Chapter 3

# Overview of the “Difference of Means” Framework

The previous chapter notes that popular indices of uneven distribution can be expressed in a variety of mathematically equivalent ways. The discussion there (and in Appendices) reviews a variety of formulas presented previously in the literature. It also introduces a set of new formulas that cast indices of uneven distribution as group differences of means on individual residential outcomes. I argue that the group difference of means formulation is an important new approach that brings many advantages and possibilities to segregation measurement and analysis. To make the case for this view I now provide a more detailed discussion comparing standard computing formulas with the difference of means formulas.

### 3.1 Index Formulas: The Current State of Affairs

As I noted briefly in the previous chapter, popular measures of segregation such as the widely used dissimilarity or delta index (D) traditionally have been formulated and interpreted from a perspective that focuses attention on outcomes for areas rather than outcomes for individuals. For example, the following formula from Duncan and Duncan (1955: 211) highlights area differences in relative group presence – specifically, the area’s share (s) of the group’s city-wide population – for the two groups in the comparison. This formula is widely used to compute D because it is computationally efficient and easy to implement. In addition, the focus on variation in area outcomes is seen as an appealing basis for understanding and assessing the extent to which two groups are distributed unevenly across the residential areas of a city.

$$D = 100 \cdot \frac{1}{2} \sum \left| \left( \frac{n_{1i}}{N_1} \right) - \left( \frac{n_{2i}}{N_2} \right) \right|$$

$$D = 100 \cdot \frac{1}{2} \sum |s_{1i} - s_{2i}|$$

This aggregate-level approach is not unique to the era in which Duncan and Duncan were writing or to the dissimilarity index. More than four decades later Hutchens introduced a new measure of uneven distribution termed the square root index (R) and drew on a similar formulation to clarify how R assesses the extent to which two groups are distributed unevenly across the residential areas of a city (Hutchens 2001: 23).

$$R = 100 \cdot \left( 1.0 - \sum \sqrt{(n_{1i} / N_1) \cdot (n_{2i} / N_2)} \right)$$

$$R = 100 \cdot \left( 1.0 - \sum \sqrt{s_{1i} \cdot s_{2i}} \right)$$

In the formula for D, uneven distribution is assessed as 0 only when the “area share scores” (s) for the two groups in the comparison are exactly equal in all areas of the city. The same is true for the formula for R.<sup>1</sup>

Summary measures of uneven distribution formulated in this way have been and remain valuable tools for aggregate-level description. But the focus on outcomes for areas rather than individuals and groups imposes a significant limitation that Duncan and Duncan (1955) noted over 50 years ago. The limitation is that area-oriented formulations of D and other indices provide little basis for gaining insight into how underlying micro-level social processes of residential attainment give rise to the area patterns that determine the level of residential segregation for the city. Accordingly, Duncan and Duncan stated “In none of the literature on segregation indices is there a suggestion about how to use them to study the *process* of segregation or change in the segregation pattern” (1955: 223; emphasis in original). The process of course plays out at the level of individuals and households, not for areas. Indeed, the areas often are defined as statistical units with no intrinsic sociological qualities relevant for segregation process; they are merely useful constructs for assessing group differences in residential distribution. So formulas that focus attention on outcomes for areas are at a level of abstraction removed from “where the action is” in segregation dynamics.

Duncan and Duncan additionally noted it would be desirable, but was not then possible, to incorporate controls for the role of individual-level factors (e.g., labor force status, occupation, income, etc.) beyond race when seeking to understand and explain the level of segregation in a city. Unfortunately, efforts to achieve this goal were frustrated then and are currently frustrated now by thinking about segregation solely from the point of view of the area-oriented computing formulas given above. The formulas are framed in terms of outcomes for areas, not in terms of individual

---

<sup>1</sup> This may be less obvious for R because it is a relatively recent addition to the literature. When share scores (s) for groups are equal in an area, the square root of the product of the two share scores will equal the value of the individual share scores. The resulting terms will sum to 1 when group share scores are equal in all areas and the index score will be 0.

residential outcomes. So it is no surprise that it is not easy to use them to gain insights into how index scores arise from an underlying micro-level process where potentially many factors play a role in shaping the residential outcomes individuals attain.

When segregation when conceptualized and analyzed from the point of view of outcomes for areas, it is very difficult to take account of the role of even a single social or economic characteristic beyond race and it is completely infeasible to take account of the role of several social and economic characteristics at the same time. Past efforts to achieve the goal of controlling for the role of non-racial characteristics have been limited to computing index scores using group subsamples that are matched on one or more relevant social characteristics (e.g., income). This approach is untenable in practical application because analysis quickly comes to be based on very small subgroup counts if one measures non-racial characteristics in fine-grained ways and/or if one tries to control for more than one or two non-racial characteristics at the same time. Accordingly, the approach is used infrequently in the empirical literature. When it is used, implementations are crude and unsatisfying and the resulting index scores are likely to be problematic on technical grounds. The implementations are crude because fine-grained distinctions quickly lead to small subgroup counts. Consequently, “matching” on non-racial characteristics can at most involve one or two characteristics and an interval variable such as income must be grouped into very broad categories. Yet even with these compromises, subgroup counts wind up being much smaller than overall counts and this then leads to technical problems relating to index bias, a concern I discuss in detail in Chaps. 14, 15, and 16.

In short, it is a disappointing state of affairs. In the six decades that have passed since Duncan and Duncan raised these important and fundamental concerns, the problems they identified have yet to be adequately addressed. Researchers continue to formulate indices of uneven distribution from area-oriented perspectives that leave the connections between index scores and individual-level residential attainments, and the related micro-level processes that shape them, unspecified and poorly understood. As a consequence, research on residential segregation has become increasingly out of step with the broader literatures investigating racial and ethnic inequality and disparity in socioeconomic outcomes such as education, occupation, and income. Studies of racial and ethnic differences in other socioeconomic outcomes have for many decades routinely drawn on micro-level models of individual attainment to gain insights into how many different factors may contribute to the creation of aggregate-level (i.e., national- and community-level) group disparities. In contrast, the literature on segregation has had to limit its focus to assessing aggregate-level segregation leaving the implications for and connections to group differences in individual residential outcomes uncertain and unexamined.

To be fair, a vibrant and important literature focusing on individual-level residential attainment has emerged in recent decades (e.g., Alba and Logan 1993; Logan and Alba 1993; Logan et al. 1996; Alba et al. 1999; South and Crowder 1997, 1998). But it has developed as a separate literature that is only loosely connected with research investigating segregation at the aggregate-level. The reason for this is that

the dependent variables in analyses of individual residential attainment do not correspond to terms that figure directly in the calculation of segregation index scores. Accordingly, studies of individual residential attainments to date do not, and logically cannot, provide direct insights into the values of D or other aggregate-level summary indices of uneven distribution. Conversely, studies of aggregate-level segregation cannot directly provide insights into the parameters of individual-level residential attainment processes.

The current state of affairs is unfortunate and unsatisfactory. Interest in segregation generally rests on an implicit assumption that segregation has important associations with group differences on neighborhood residential outcomes that are relevant for socioeconomic attainment and inequality in life chances. Individuals and households strive to attain these residential outcomes either for their own sake (e.g., as markers of social position) or because they are closely correlated with factors that impact life chances (e.g., exposure to crime, social problems, schools, services, neighborhood amenities, etc.). In view of this, it is clearly desirable to gain a better understanding of how different segregation indices relate to group differences on individual-level residential outcomes. Surprisingly, the methodological literature on segregation measurement is nearly silent on this issue. Segregation measurement theory gives attention to many properties and qualities of aggregate-level indices but it has not taken up the question of how different indices relate to individual-level residential outcomes or carry different implications for group differences on residential outcomes.

### 3.2 The Difference of Means Formulation – The General Approach

I address this gap in the measurement literature by casting popular measures of uneven distribution as differences of group means on segregation-relevant individual residential outcomes. Specifically, I place familiar segregation indices in a common “difference of means” framework in which the index score “S” is given as

$$S = Y_1 - Y_2$$

where:

S is the score of the relevant segregation index (i.e., G, D, R, H, or S),

$Y_1$  is the mean on y for individuals in Group 1 based on either  $(1/N_1) \cdot \sum n_{1i} y_i$  when computed for area data or  $(1/N_1) \cdot \sum y_{1j}$  when computed for individual data,

$Y_2$  is the mean on y for individuals in Group 2 based on either  $(1/N_2) \cdot \sum n_{2i} y_i$  when computed for area data or  $(1/N_2) \cdot \sum y_{2j}$  when computed for individual data,

$n_{1i}$  and  $n_{2i}$  are the counts of Groups 1 and 2, respectively, in the i'th area,

$p_i$  is the pairwise area proportion for Group 1 in the i'th area based on  $p_i = n_{1i} / (n_{1i} + n_{2i})$ ,

$y_i$  is the residential outcome score ( $y$ ) for the  $i$ 'th area scored as a function of the pairwise area group proportion  $y_i = f(p_i)$ ,  
 $y_{1k}$  indicates the residential outcome ( $y$ ) for the  $k$ 'th individual in Group 1 (set equal to the residential outcome score for the area in which the individual resides), and  
 $y_{2k}$  indicates the residential outcome ( $y$ ) for the  $k$ 'th individual in Group 2 (set equal to the residential outcome score for the area in which the individual resides).

I hold that formulating segregation indices in this way is useful for both conceptual and practical reasons. First, it provides a new interpretation for aggregate segregation indices; they now can be understood as registering simple group differences on residential outcomes ( $y$ ) scored based on area group proportion ( $p$ ) which has an easy, straightforward interpretation as (pairwise) contact with or exposure to Group 1 (i.e., the reference group) based on co-residence. Simple co-residence, of course, does not necessarily imply harmonious social interaction. But it does indicate common fate regarding many neighborhood outcomes and many shared residential experiences. On this basis, it is a potentially important and meaningful social indicator.

Second, this new approach to computing index values places different indices in a uniform, common computing framework that highlights differences between measures on a single, specific point of comparison – the manner in which each index registers neighborhood residential outcomes ( $y$ ) based on area group proportion ( $p$ ). Since area group proportion can be understood as contact or exposure based on co-residence with Group 1, all of the indices can be interpreted as group differences in average “scaled contact” with Group 1. Differences between indices ultimately trace to differences in the specific way that residential outcomes ( $y$ ) are quantitatively scored based on area group proportion ( $p$ ). Consequently, differences between indices can be seen as arising solely from differences in the index-specific form of the scaling function  $y=f(p)$ . This provides a new basis for evaluating segregation indices; they can be compared on the substantive relevance of how each index registers residential outcomes ( $y$ ) based on contact and exposure with Group 1 as embodied by area group proportion ( $p$ ).

Third, the segregation-relevant residential outcomes ( $y$ ) used to compute the segregation index score can directly serve as dependent variables in individual-level residential attainment analyses. Thus, in the difference of means formulation, the segregation index score can be equated to the effect of group membership (e.g., coded 0 or 1) in an individual-level residential attainment analysis for the city. This carries minimal practical value for specific task of estimating index scores because the scores can be readily obtained by simpler methods. But it is important because it expands options for understanding and analyzing segregation. It unifies the study of aggregate segregation with the study of residential attainment in a single framework. In doing so it opens the door to a host of new options for segregation analysis including, for example, the ability to easily take account of the role that factors other than group membership (e.g., income) may play in determining segregation and the ability to use multi-level models of residential attainment to study cross-area and cross-time variation in segregation.

### 3.3 Additional Preliminary Remarks on Implementation

The key to implementing the new approach is to identify for each index a scoring system for neighborhood outcomes ( $y$ ) that will yield the segregation index score as a difference of group means ( $Y_1 - Y_2$ ). I have identified relevant scoring systems for five indices that are widely used to measure the unevenness dimension: the gini index (G), the delta or dissimilarity index (D), the separation index (S) (also known as the variance ratio index [V]), the Theil entropy index (H), and the Hutchens square root index (R), a measure that is closely associated with the “symmetric” implementation of the Atkinson index (A).<sup>2,3</sup>

For all of these indices, the residential outcome ( $y$ ) is scored as a function of “pairwise” group proportion ( $p$ ) for the area the individual resides in. Indexing areas by “ $i$ ”,  $p_i$  is given as

$$p_i = n_{1i} / (n_{1i} + n_{2i})$$

where  $p_i$  is the Group 1 proportion in the combined population of Group 1 and Group 2 in the  $i$ 'th area. In this formulation the scoring system for each index rests on a “scaling” function  $y=f(p)$  that maps area group proportion scores ( $p$ ) on to index-specific residential outcome scores ( $y$ ). I discuss the index-specific scaling functions  $y=f(p)$  in the chapters that follow and in Appendices.

Before continuing, I comment briefly to note two technical points. One is that the designation of which group serves as “Group 1” is arbitrary. One group must be so designated. But the result for the index score will be the same regardless of which of the two groups is chosen as the reference. White (1986) termed this index property as “symmetry.” When one group is understood as a majority group and the other as a minority group, it has been conventional in previous research to designate the majority group as Group 1. This is not required, but it is convenient because it facilitates interpreting segregation as reflecting the extent to which the minority group has less contact with the majority group than would occur under even distribution. This has generally been viewed as useful based on the assumption that areas of majority group residence are advantaged and thus disparity and disadvantage in residential outcomes follows when contact with the majority falls below parity. But it is only a custom, not a logical requirement. If the roles of the two groups are

---

<sup>2</sup>The separation index (S) is known by many names including the revised index of isolation (Bell 1954), the correlation ratio and eta squared ( $\eta^2$ ) (Duncan and Duncan 1955),  $r$  or  $r_{ij}$  (Coleman et al. 1975, 1982), variance ratio (V) (James and Taeuber 1985), and segregation index (S) (Zoloth 1976).

<sup>3</sup>I note below that A is an exact nonlinear function of R. To date I have not identified the relevant scoring system for Atkinson’s Index (A). A is rarely used in empirical studies and has been criticized on conceptual grounds for being asymmetric such that A for White-Black segregation may differ from A for Black-White segregation (e.g., White 1986). But it has received attention in the segregation measurement literature. So this remains a task for future research.

reversed, the contact interpretation will be reversed. But all substantive implications of the patterns of group differences in contact will remain intact and unchanged.

The second technical point I mention is that  $p$  is computed using only counts for the two groups in the segregation comparison. Thus,  $p$  is not Group 1's proportion among the *total* population of the area; it is Group 1's proportion among the combined count of the two groups in the segregation analysis. To emphasize this point, I sometimes term  $p$  as a "pairwise" group proportion. However, as this is the primary way I use  $p$  in this monograph, I often drop the "pairwise" modifier in the interest of economy of expression. Note that this "pairwise" construction is not at all controversial in segregation measurement; relevant terms in all of the standard formulas for measures of uneven distribution reviewed earlier are based on pairwise implementations of group proportions for areas (i.e.,  $p$  and  $q$ ) and for the city as a whole (i.e.,  $P$  and  $Q$ ).

The general outline of the approach is now set. The next task is to review how the difference of means framework can be implemented with the most popular and widely used segregation indices.

## References

- Alba, R. D., & Logan, J. R. (1993, May). Minority proximity to whites in suburbs: An individual-level analysis of segregation. *American Journal of Sociology*, *98*, 1388–1427.
- Alba, R. D., Logan, J. R., Stults, B., Marzan, G., & Zhang, W. (1999, June). Immigrant groups and suburbs: A reexamination of suburbanization and spatial assimilation. *American Sociological Review*, *64*, 446–460.
- Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces*, *32*, 357–364.
- Coleman, J. S., Kelly, S. D., & Moore, J. A. (1975). *Trends in school segregation, 1968–1973*. Washington, DC: The Urban Institute.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). Achievement and segregation in secondary schools: A further look at public and private school differences. *Sociology of Education*, *55*, 162–182.
- Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indices. *American Sociological Review*, *20*, 210–217.
- Hutchens, R. (2001). Numerical measures of segregation and their desirable properties. *Mathematical Social Sciences*, *42*, 13–29.
- James, D., & Taeuber, K. (1985). Measures of segregation. *Sociological Methodology*, *13*, 1–32.
- Logan, J. R., & Alba, R. D. (1993, May). Locational returns to human capital: Minority access to suburban community resources. *Demography*, *30*, 243–268.
- Logan, J. R., Alba, R. D., McNulty, T., & Fisher, B. (1996). Making a place in the metropolis: Locational attainment in cities and suburbs. *Demography*, *33*, 443–453.
- South, S. J., & Crowder, K. D. (1997, November). Residential mobility between cities and suburbs: Race, suburbanization, and back-to-the-city moves. *Demography*, *34*, 525–538.
- South, S. J., & Crowder, K. D. (1998, February). Leaving the 'Hood: Residential mobility between black, white, and integrated neighborhoods. *American Sociological Review*, *63*, 17–26.

- White, M. J. (1986). Segregation and diversity: Measures of population distribution. *Population Index*, 65, 198–221.
- Zoloth, B. S. (1976). Alternative measures of school segregation. *Land Economics*, 52, 278–298.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 2.5 International License (<http://creativecommons.org/licenses/by-nc/2.5/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

