Measures of Nominal and Ordinal Population Diversity

John R. Ottensmann
Indiana University-Purdue University Indianapolis
john.ottensmann@gmail.com
urbanpatternsblog.wordpress.com

Abstract

A measure of population diversity across a set of categories such as racial and ethnic groups is presented that clearly demonstrates how the index captures deviations from a standard of maximum diversity. The measure is applied to subareas of a larger area such as census tracts and is employed to develop 2 alternative measures of neighborhood diversity for the larger area. This and other measures of diversity assume that the population is classified into a nominal, unordered set of categories. The distribution of the population into a set of rank-ordered categories introduces the additional element that the pattern of a distribution across the categories should be considered in assessing levels of diversity. A measure of ordinal diversity and corresponding measures of neighborhood ordinal diversity are presented to address this.

Introduction

For a population distributed among a number of groups of interest, the question arises as to the extent to which the population is more evenly distributed among those groups as opposed to being concentrated in one or a few of the groups. The issue arose in considering the racial and ethnic composition of the populations of urban areas over time. This required a measure of the diversity of the population among the different groups.

Several indexes have been used to measure the diversity of everything from the distribution of species in an ecosystem to the concentration of economic activity among firms in an industry. While these are reasonable measures, it is not obvious how they capture the level of diversity. I chose to develop a new, closely related index, that makes clear how it is measuring the diversity of the distribution of the population across a set of categories. This is the first measure, of area nominal diversity, presented in this paper.

Of course such an index could be used to measure diversity in areas of any size. A special case, however, is the measurement of diversity in small subareas such as census tracts within a larger area. Not only is the diversity of the individual areas of interest, so is the combined level of diversity in these smaller areas within the larger
area, which could be considered to be a measure of neighborhood diversity for that larger area. But such a measure of neighborhood diversity is necessarily limited by the level of the diversity in that larger area. Therefore additional consideration is given to the level of neighborhood diversity relative to the maximum possible given the level of diversity (or lack thereof) in the larger area.

This initial exploration of the measurement of diversity assumed that the population is distributed into a set of categories established using a nominal classification. That is, the groups are not ordered. The division of the population into racial and ethnic groups is an example of such a classification. But other variables can classify the population into an ordered set of categories. The distribution of the population by level of education is one example. In such cases, the ordering of the groups matters. The concentration of the population in groups at one end of the distribution should be seen as being less diverse than a similar distribution of the population among groups across the spectrum. This suggests the use of measures of ordinal diversity that take such ranking of the groups into account. And these can also be applied to an entire area and to smaller subareas, with those values being combined to create measures of neighborhood ordinal diversity for the larger area.

**Measuring Racial and Ethnic Diversity**

This investigation into the measurement of diversity began with a focus on the measurement of racial and ethnic diversity in urban areas. This section reviews some of the efforts at measuring such diversity.\(^1\)

Numbers of studies of neighborhood racial diversity and integration have simplified the measurement by identifying 1 or more simple measures or classifications. Hero and Tolbert (1996) used 2 measures: percent nonwhite for minority diversity and percent southern and Eastern European for white ethnic diversity. Attempting to measure stable neighborhood racial integration, Ellen (1998) considered a neighborhood to be integrated if it were 10 to 50 percent black and stable with a change in non-Hispanic whites of less than 10 percent. Approaching the same problem, Galster (1998) defined a mixed-race neighborhood as one in which no more than 75 percent of the population belonged to a single racial or ethnic group. Holloway, Wright, and Ellis (2012) classified tracts as low and moderate diversity white, black, Latino, Asian, or as highly diverse and then cross classified these to create a typology of diversity. They do make reference to entropy values in defining their arbitrary cutoffs.

Others have developed single continuous measures of diversity that are used to measure the overall level of diversity in different areas. White (1988) provided a helpful

---

\(^1\) This section and the following section have been adapted from my paper on racial and ethnic diversity in urban areas (Ottensmann 2019a).
review of common segregation and diversity measures of population distributions. He identified 2 primary measures of population diversity, which he termed the entropy index and the interaction index. These measures and some applications are discussed here.

The idea of entropy as a concept that might be used as a measure of population diversity is often traced back to Shannon’s (1948) application of the idea of entropy to information theory. The index is also sometimes called Theil entropy given his contributions furthering the development of the concept leading to its greater use (Theil 1972). Entropy as a measure of population diversity can be expressed as

\[ E = - \sum_{i=1}^{n} p_i \log(p_i) \]

where \( p_i \) is the proportion of the population in each of \( n \) groups and \( \log(p_i) \) is the natural logarithm. The index has the minimum value of 0 with all of the population in a single group and increases as the diversity increases. The maximum value of the index depends on the number of (nonzero) groups and is \( \log(n) \).

The index can be normalized to range from 0 to 1 by dividing by the log of the number of groups. However this creates a problem for comparisons involving a set of groups in which some units can have zero members in a group, effectively reducing the number of groups for the entropy index, changing the value of \( n \) used in normalizing by dividing by \( \log(n) \). This is because \( \log(p_i) \) is undefined when the proportion equals zero. But then if 1 person is added to a group with formerly no members, while this has a negligible effect on the value of the sum, the normalized index value would experience a discontinuous decrease with the increase in the value of \( \log(n) \) as the divisor for normalizing the index. An alternative approach that avoids this problem would be the substitution of 0 for the product when the proportion is 0 and the log is undefined given that the value of the product approaches 0 as the limit when the proportion approaches 0. But this seems to be an unsatisfactory way around the problem.

This problem can be seen in the study by Dougherty (2003) of diversity in religious congregations and parishes. He uses the normalized entropy index as the measure of diversity, but only for the groups represented in each congregation or parish. And given the context, many of the congregations or parishes would not have had members from all of the racial and ethnic groups considered.

The index White (1988) called the interaction index was developed by multiple persons working in different fields. Simpson (1949) proposed his diversity index as a measure of species diversity in an ecosystem. Herfindahl (1950) and Hirschman (1964) developed the index as a measure of industrial concentration in economics, where it is known as the Herfindahl Hirschman Index or HHI. For simplicity, further references to
this index will be simply to the Herfindahl index, as that is the source most commonly cited in the literature that is relevant here. The measure is simply the sum of the squared proportions, in the present case in a set of groups of a population:

$$H = \sum_{i=1}^{n} p_i^2$$

where once again $p_i$ is the proportion of the population in each of $n$ groups. The index is the probability that 2 members of the population selected at random would belong to the same group (technically with selection with replacement or from an infinite population). The index takes on a maximum value of 1 when the entire population is concentrated in a single group. The minimum value comes with equal proportions in each group. This depends on the number of groups $n$ and is simply $1/n$. Thus the index increases as diversity decreases and might more appropriately be seen as a measure of concentration. While the index could be normalized to range from 0 to 1, this is not often seen.

Schmid, Al Ramiah, and Hewstone 2014 used the Herfindahl index in a study of the relationship of neighborhood ethnic diversity to trust in Britain. However, they apparently used only 2 groups, white and minority so they did not benefit from the full value of using the index for larger numbers of groups. Others have used 1 minus the Herfindahl index to give a measure that increases as diversity increases. Alisena and La Ferrara (2005) reviewed literature on economic effects of various kinds of diversity across nations and cities and said that most studies used this measure, which they termed the “fractionalization” index. Graif and Sampson (2009) measured language diversity in this way. Seaton and Yip (2009) likewise used 1 minus the Herfindahl index to measure racial diversity in both neighborhoods and schools. Using an alternative approach, Talen (2005) used the inverse of the Simpson/Herfindahl index to examine multiple types of diversity at the neighborhood level.

**Area Nominal Diversity**

This section presents a measure of the overall diversity for an entire area (of any size) that was first developed for my research on racial and ethnic diversity in urban areas (Ottensmann 2019a). This is being referred to as area nominal diversity as it assumes that the population is distributed across a number of groups with the classification being a nominal variable. The groups have no inherent ordering, which is certainly the case with a racial and ethnic classification.

An ideal index of diversity would be transparent and intuitive. It should be obvious as to how it is measuring diversity. The index should have a maximum value
when equal proportions are in each of the groups and a minimum value when the entire population is concentrated in one group. It should be normalized to range from 0 to 1 (or 100). And the index should not have problems when there are no persons in one or more groups. (This last criterion would not be necessary for the measurement of diversity in large urban areas, as all have at least some persons in every group. However the index should also be applicable to the measurement of neighborhood diversity, where many census tracts will not have representation of every group.)

The Herfindahl index meets or can be modified to meet all of the criteria except for transparency. It is not immediately obvious how the sum of squared proportions serves as a measure of diversity. To be sure, some simple, quick numerical examples can illustrate how it works. And consideration of how changes in the squares of numbers less than 1 vary as the numbers become smaller might do the same. But these do not equate with transparency.

Creation of the index of diversity to be used here begins with the assumption that diversity is greatest with equal proportions of the population in each of the groups. White (1988) and many others have asserted this principal as well. Then differences between the proportions actually in each group and the equal proportions value represent departures from maximum diversity. Squaring these differences and summing them gives an overall measure of the total departure from complete diversity:

$$\sum_{i=1}^{n} \left( p_i - \frac{1}{n} \right)^2$$

where again $p_i$ is the proportion of the population in each of $n$ groups. This sum is not a measure of diversity but rather of the departure from maximum diversity and the concentration of the population in some of the groups. The sum has a minimum value of zero when all of the proportions are equal, maximum diversity. The maximum value is $1 - 1/n$ when the entire population is concentrated in a single group, minimum diversity.

To create the index of diversity with the desired properties, the sum is first multiplied by the inverse of its maximum value of $1 - 1/n$ to normalize it to range from 0 to 1. This is then subtracted from 1 so the resulting index ranges from 0 for minimum diversity, concentration in a single group, to 1 for maximum diversity, equal proportions in each group:

$$D = 1 - \frac{n}{n-1} \sum_{i=1}^{n} \left( p_i - \frac{1}{n} \right)^2$$
While this is a perfectly reasonable formulation for the diversity index, using simple algebra the formula can be re-expressed in the following form that is it slightly more convenient for computation and provides greater numerical accuracy:

\[ D = 1 - \frac{n}{n-1} \left( \sum_{i=1}^{n} p_i^2 - \frac{1}{n} \right) \]

Note that in the presentation of the diversity index, it may be reasonable to multiply the index by 100 so the index ranges from 0 to 100. This makes tables and individual values easier to read as they are not all prefixed by the zero followed by the decimal point.

If this is starting to look a lot like the formulas for the variation, variance and standard deviation, this is no accident. The sum of squares for the variation (then divided by \(n\) or \(n - 1\) to give the variance) is the sum of the squared differences between the values of the variable and the mean. The index of diversity starts with the sum of the squared differences between the proportions and \(1/n\). But since the sum of the proportions is 1, \(1/n\) is the mean proportion. The variance also has a computing formula including the sum of the squared values. Subtracted from this is the squared sum of the values divided by \(n\). But again since the sum of the proportions equals 1, the sum of the squared values is also equal 1, giving the \(1/n\) term in the computing formula for the diversity index. The diversity index can therefore be described as the variation in the proportions, normalized to range from 0 to 1, and subtracted from 1 to increase as diversity increases and the variation decreases.

The final, computing formula for the index of diversity also suggests a relationship to the Herfindahl index. The Herfindahl index sum of squared proportions is literally embedded in the computing formula. But the relationship is much closer. The maximum value for the Herfindahl index is 1 and the minimum value is \(1/n\). To normalize the Herfindahl index to range from 0 to 1, one must first subtract \(1/n\) and then multiply that by the inverse of \(1 - 1/n\). Subtract that from 1 to reverse the direction and one gets the computational formula for the diversity index, \(D\). In other words, the index of nominal diversity developed here is equivalent to the Herfindahl index normalized to range from 0 to 1 and subtracted from 1.

Table 1 gives examples of the nominal diversity \(D\) for several distributions of the population among 4 groups. It gives the proportions in each of the groups and the value for the nominal diversity index. The minimum value is associated with the concentration of the entire population in 1 group, giving a diversity of 0. The maximum value comes with the uniform distribution of the population across all 4 groups, producing the value of 1. For half of the population in 2 of the groups and none in the other 2, the index is 0.67 and for a third in 3 of the groups the index is 0.89.
This index of nominal diversity has been used in 2 papers examining the racial and ethnic diversity in 59 large urban areas from 1980 to 2010 (Ottensmann 2019a, 2019b). It has performed well, showing a wide range of variation across the urban areas at the different times. For the urban areas as a whole, the minimum level of diversity for any urban area was 14 for one area in 1980, and the maximum was 92 for an area in 2010.

Subarea (Neighborhood) Nominal Diversity

The index of nominal diversity D presented in the following section may, of course, be used as a measure of diversity in areas of any size. Here we are interested in the case in which diversity is initially measured in the smaller subareas of a larger area. A common application would be to the measurement of diversity in the census tracts within an urban area. In such a case this might be called neighborhood nominal diversity.

The formula for such subarea nominal diversity is essentially the same as presented above, with the addition of a subscript to identify the census tract or other subarea for which the diversity is being computed:

\[ N_j = 1 - \frac{n}{n - 1} \sum_{i=1}^{n} \left( p_{ij} - \frac{1}{n} \right)^2 \]

where \( N_j \) is the neighborhood nominal diversity for census tract (or other subarea) \( j \) and \( p_{ij} \) is the proportion of the population of census tract \( j \) in group \( i \). The individual measures of tract diversity may be of interest themselves, for example to look at the range of levels of neighborhood diversity within the urban area or to examine the relationship of diversity to other tract characteristics. It would have hardly been worth

Table 1. Nominal Diversity for a Sample of Distributions across 4 Groups.

<table>
<thead>
<tr>
<th>Proportional Distribution</th>
<th>Nominal Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>
repeating the formula just to add the tract subscripts for such purposes. But this measure will also be used to develop measures of overall neighborhood diversity for the entire urban area.

Continuing with the example of an urban area subdivided into census tracts, one would like to have a measure of neighborhood (tract) diversity for the urban area as a whole. An obvious measure would be the average of the tract diversities. Since census tracts vary in population, the mean should be weighted by the populations of the tracts. The (first) measure of urban area neighborhood diversity is then:

\[ N^A = \frac{1}{\sum_{j=1}^{m} T_j} \sum_{j=1}^{m} T_j N_j \]

where \( N^A \) is the absolute neighborhood nominal diversity for the entire area and \( T_j \) is the total population in census tract \( j \) for the \( m \) tracts in the urban area. This measure is being called absolute diversity and is denoted with the superscript \( A \) as it is the direct average of the levels of diversity in the census tracts.

The absolute neighborhood diversity \( N^A \) will have a minimum value of 0 when every tract is occupied by members of a single group. A maximum value for absolute neighborhood diversity of 1 can only be reached if the diversity in each tract is also 1. This would be the case if the proportions in each group were equal in every tract. But this could occur only when the proportions in each group are equal in the entire urban area, giving a value for urban area diversity \( D \) also of 1. To the extent the distribution among the groups is not equal for the urban area as a whole and the urban area diversity \( D \) is less than 1, the maximum possible value for absolute neighborhood diversity \( N^A \) will likewise be less than 1.

The index of dissimilarity, the most commonly employed measure of segregation (Duncan and Duncan 1955), reaches the maximum value of 1 when each tract is occupied by members of one the the groups. The minimum value of 0, no segregation, is reached when the distributions of the members of the 2 groups across the census tracts are the same or, equivalently, when the proportions in each group are the same for each tract. This is considered to be the absence of segregation. It is not dependent on the proportions in each group within the urban area.

The maximum value for the absolute neighborhood diversity index \( N^A \) is dependent on the distributions of the urban area population among the various groups. It is desirable to have an alternative measure of neighborhood diversity for the urban area that ranges from 0 to 1 with 1 representing the maximum neighborhood diversity possible given the urban area racial and ethnic distribution. This would be the equivalent of the absence of segregation for more than 2 groups.
The maximum value for the absolute neighborhood diversity index occurs when all of the census tracts have identical distributions of the members of the groups. This is shown as follows: Start with that case, identical distributions in every tract. Now change the distribution in the first tract by increasing slightly the proportion in 1 of the groups, while correspondingly decreasing the proportion in a second group. In a second tract, the proportion in the first group will have to be decreased to account for the numbers contributing to the increase in the first tract and the numbers in the second will have to be increased. So the second tract would also have increases and decreases in the proportions. The proportions are squared for the summation for the tract neighborhood diversity index. The increased proportions in both tracts will contribute to an increase in their sums. And of course the decreased proportions will make the sums smaller. But increasing a value between 0 and 1 by a set amount produces a larger change in the squared value than decreasing a value by the same amount and squaring it. So the sums of the squared proportions will be larger in both tracts and since the index uses the negative value of the sum, the tract neighborhood diversities will be smaller for both tracts. The absolute neighborhood diversity for the urban area, being the weighted mean of the tract neighborhood diversities, will thus be smaller. Therefore, the case in which the distributions of the groups is identical for all of the tracts is the maximum possible value for absolute neighborhood diversity. And this is comparable to the absence of segregation with an index of dissimilarity value of 0 for 2 groups.

Now if the distributions of the groups are identical for all of the tracts, the tracts will have equal tract neighborhood diversity values and the weighted average for all tracts will produce the same value for absolute neighborhood diversity. And if the distributions are the same for all tracts, this must be the distribution across the groups for the urban area as a whole. So the diversity index $D$ for the urban area will be the same as the tract neighborhood diversities for all of the tract. So it follows that the maximum value for absolute neighborhood diversity is the level of diversity for the urban area. The more diverse the urban area as a whole, the greater the possible values for absolute neighborhood diversity, which is reasonable and intuitive. Greater diversity implies larger proportions for the smaller groups, allowing larger proportions in the tracts and greater tract neighborhood diversity.

This forms the basis for the second measure of neighborhood diversity for the urban area. It is produced by rescaling the absolute neighborhood diversity index to have a maximum value of 1 for the maximum possible value for the index for that urban area. This new index is then

$$N^R = \frac{N^A}{D}$$
where $N_R$ is the relative neighborhood diversity index for the urban area, the absolute neighborhood diversity divided by its maximum possible value for the urban area, overall urban area diversity $D$. This index is more comparable to the index of dissimilarity for segregation, as it can span than the full range from 0 to 1 regardless of the distribution among the groups and the diversity for the urban area. The direction is reversed, however, with maximum segregation having a value of 0 for relative neighborhood diversity while it is 1 for the index of dissimilarity. The absolute neighborhood diversity index, on the other hand, is a measure of the average level of diversity at the tract level, period, without consideration of limitations on that diversity imposed by the level of diversity or lack thereof in the urban area overall.

**Area Ordinal Diversity**

The measures of diversity discussed to this point and used for analyses of racial and ethnic diversity assumed a nominal classification of the population. There was no ordering among the groups. They were therefore called measures of nominal diversity.

Other classifications place the population into a set of ordered or ranked groups, Education would be one example. The population might be divided into 4 groups, those who have not graduated from high school, high school graduates, those with some college, and finally those with a bachelor’s degree or higher. For considering diversity, such ordering matters. A population with half high school graduates and half college graduates could reasonably be considered to be somewhat more diverse than one with half high school graduates and half who had not graduated from high school. To capture such differences requires a measure of ordinal diversity that takes into account the ordering of the groups.

A measure of ordinal diversity is created that combines 2 components. The first captures the diversity of the population considering the distribution among the groups without taking into account of the ordering of the groups. This is the nominal diversity component. The second component adds the effect associated with the pattern of the distribution taking into account the ordering of the groups. This is the ordinal diversity component.

The nominal component is that used in the measure of nominal diversity. Deviations from maximum diversity—equal proportions in the groups—are squared and then normalized to range from 0 to 1:

$$P = \frac{n}{n-1} \sum_{i=1}^{n} \left( p_i - \frac{1}{n} \right)^2$$
The ordinal component must consider deviations from a distribution that reflects the ordering of the groups. Cumulative distributions have this property, with each group including the quantities of all of the preceding groups. So we take as the standard for computing ordinal diversity the cumulative distributions derived from the distribution with equal proportions in each group. And we then take the differences between the proportion in each groups and the cumulative distributions, square those, and sum them. Because of the increasing values of the cumulative distribution, the differences will generally become greater as one approaches the top of the cumulative distribution. Of course this captures the greater effect of higher proportions in the more extreme groups in only one direction. So it is necessary to use 2 cumulative distributions across the uniform maximum diversity distribution, one in each direction, taking the squared deviations from each and summing those separately. These are the two sums for the differences from the cumulative distribution up and the cumulative distribution down:

\[
\sum_{i=1}^{n} \left[ p_i - i \left( \frac{1}{n} \right) \right]^2
\]

\[
\sum_{i=1}^{n} \left[ p_i - (n + 1 - i) \left( \frac{1}{n} \right) \right]^2
\]

The second term in the brackets represents the cumulative distributions of the maximum diversity equal proportion distribution from the first group to the last and then the reverse.

These terms capture the extent to which the distribution in weighted in 1 direction or the other. The extent to which the distribution is unevenly weighted with respect to the ordered groups is then captured by the difference between the sums:

\[
\text{abs} \left\{ \sum_{i=1}^{n} \left[ p_i - i \left( \frac{1}{n} \right) \right]^2 - \sum_{i=1}^{n} \left[ p_i - (n + 1 - i) \left( \frac{1}{n} \right) \right]^2 \right\}
\]

The absolute value of the differences is required as the difference can be positive or negative.

The maximum value for this expression, which occurs when the entire population is concentrated in either the bottom or the top group, is greater than 1. This maximum is needed to normalize the expression to vary between 0 and 1.
Determining the value if the proportion in the first group is 1—the maximum diversity possible—is made easy by recognizing that for both sums, the differences for the sets of terms except the first and last will be the same, 0 minus the values for the cumulative distributions. They will just occur in the opposite order. So it is only necessary to consider the values for the first and last term.

For the first summation with differences from the cumulative distribution starting with the first term, the value of the first term for the case with the proportion for that term being 1 is

\[
(1 - \frac{1}{n})^2
\]

And the value for the last term is 1, the difference between the proportion 0 and the final value of the cumulative distribution, which is 1 which is of course also the value squared. For the second summation the first value is 0, the proportion 1 minus the final value of the cumulative distribution down which is 1. The value of the last term is simply

\[
\left( -\frac{1}{n} \right)^2 = \frac{1}{n^2}
\]

So the maximum difference between the the sums is

\[
\left( 1 - \frac{1}{n} \right)^2 + 1 - \frac{1}{n^2} = 2 \left( \frac{n - 1}{n} \right)
\]

We then multiply the absolute value of the differences by the inverse of the maximum value to normalize that difference to range from 0 to 1:

\[
Q = \frac{1}{2} \left( \frac{n}{n - 1} \right) \text{abs} \left\{ \sum_{i=1}^{n} \left[ p_i - i \left( \frac{1}{n} \right) \right]^2 - \sum_{i=1}^{n} \left[ p_i - (n + 1 - i) \left( \frac{1}{n} \right) \right]^2 \right\}
\]

This is then the ordinal component of diversity. It has a minimum value of 0 when the distribution of the population proportions is symmetrical and 1 when the entire population is concentrated in either the bottom or top group.
The measure of ordinal diversity is then the average of the nominal and ordinal components, subtracted from 1 so that it increases from 0 to 1 as the level of diversity goes from minimum diversity to maximum diversity:

\[ R = 1 - \frac{(P + Q)}{2} \]

where \( R \) is the index of ordinal or rank-order diversity.

Table 2 presents examples of the ordinal diversity for a number of distributions of the population among 4 groups in various patterns. It gives the proportions in each of the groups and the values of the ordinal diversity index \( R \) and also the nominal diversity index \( D \) for comparison of the effects of including the ordering of the groups in the ordinal index. The minimum value of 0 for ordinal diversity occurs when the entire population is concentrated in either the bottom or top group. Note that when the population is concentrated in the second group, diversity increases to 0.33. The maximum value of 1 for both nominal and ordinal diversity occurs with the population uniformly distributed across the 4 groups. The intermediate cases are for various distributions of the population across groups, across 2 groups with half in each, and across 3 groups, with one-third in each. Of course the values of the nominal diversity are the same for different patterns of distributions across 2 groups, with an index value of 0.67, and across 3 groups, with an index value of 0.89. But the values for the ordinal diversity vary with the pattern of the distributions and can be higher or lower than the nominal diversity. The lowest values for the ordinal diversity come with the population concentrated in the groups at one end of the distribution. Higher values were associated with more spread-out distributions, with the maximum values for the population distributed among 2 groups coming with the distributions that were symmetric. (A symmetric distribution is not possible with the population distributed among 3 of the 4 groups.)

Subarea Ordinal Diversity

Following the example of neighborhood nominal diversity, ordinal diversity can be calculated for individual census tracts or other small areas within the larger area. The only difference is the addition of the subscript \( j \) to all of the previous expressions to denote the census tract, for example

\[ S_j = 1 - \frac{(P_j + Q_j)}{2} \]
Table 2. Nominal and Ordinal Diversity for a Sample of Distributions across 4 Groups.

<table>
<thead>
<tr>
<th>Proportional Distribution</th>
<th>Nominal Diversity</th>
<th>Ordinal Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.333</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>0.333</td>
<td>0.000</td>
<td>0.333</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Once again, an absolute measure of neighborhood ordinal diversity for the urban area as a whole is the weighted average of the tract ordinal diversity:

\[ S^A = \frac{1}{\sum_{j=1}^{m} T_j} \sum_{j=1}^{m} T_j S_j \]

As with absolute nominal neighborhood density, this can take a maximum value of 1 only with maximum diversity—an even distribution across the groups—in each of the census tracts. For an urban area with a lower level of ordinal diversity, the maximum possible neighborhood diversity will be less than 1.

Continuing as with neighborhood nominal diversity, we can create a measure of relative ordinal diversity which would be the absolute ordinal neighborhood diversity divided by the urban area ordinal diversity:

\[ S^R = \frac{S^A}{R} \]

However unlike the case of neighborhood nominal ordinal diversity, I have not been able to demonstrate that the maximum value of absolute neighborhood ordinal
diversity $S^A$ is the value of ordinal diversity for the larger area, $R$. Thus it may be possible that relative neighborhood ordinal diversity $S^R$ could exceed 1. However I doubt that this is likely, as it would require levels of diversity in many of the tracts to be close to or greater than the level of ordinal diversity in the larger area.

References


