Measures of Segregation and Isolation

Benjamin Forest © 2005
Dartmouth College

Please note that the following examples treat White, Black and Asian as mutually exclusive categories. In Census data, Hispanics/Latinos can be of any race.

Index of Dissimilarity (D)

The Index of Dissimilarity is the most common measure of segregation. Although it has limitations, it is relatively easy to calculate and to interpret.

The Index of Dissimilarity for two groups, Whites and Blacks, in a particular city:

$$D = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{w_i}{W_T} - \frac{b_i}{B_T}\right|$$

Where:

n = number of tracts or spatial units
w_i = number of Whites in tract i
W_T = total number of Whites in the city
b_i = number of Blacks in tract i
B_T = total number of Blacks in the city

Tract	White	Black	Asian	wi/WT	bi/BT	ai/AT	Absol. Value (wi/WT-bi/BT)	Absol. Value (wi/WT-ai/AT)	Absol. Value (bi/BT-ai/AT)
1	100	0	10	0.23	0.00	0.08	0.23	0.15	0.08
2	120	10	20	0.27	0.07	0.16	0.20	0.11	0.09
3	40	0	5	0.09	0.00	0.04	0.09	0.05	0.04
4	50	20	50	0.11	0.14	0.40	0.03	0.29	0.26
5	100	20	20	0.23	0.14	0.16	0.08	0.07	0.02
6	30	90	20	0.07	0.64	0.16	0.57	0.09	0.48
Sum	440	140	125				1.21	0.76	0.97
						D =	0.60	0.38	0.48

- Value of D represents the proportion of a group that would need to move in order to create a uniform distribution of population. In this example, 60% of Blacks (or Whites) would need to move in order achieve a uniform distribution of population by race.

- The value of D is a maximum when each tract contains only one group; it is minimized (0) when the proportion of each group in each tract is the same as the proportion in the population as a whole.

Interaction or Exposure Index (B)

There are also a number of indices that try to assess the sociological effects of segregation. Strictly speaking, they are not measures of segregation, but of **isolation**. Most provide some measure of the probability that a member of one group will meet or interact with a member of another group. There are many variations - this is one of the simplest.

For Whites and Blacks, for example,

$$B_{bw} = \sum \left({}^{n_{ib}}\!/_{N_b} \right) \left({}^{n_{iw}}\!/_{n_i} \right)$$

Where:

n_{ib} = number of Blacks in the tract
n_{iw} = number of Whites in the tract
N_b = number of Blacks in the city
n_i = total population of the tract

Tract	White	Black	Asian	n_{ib}/N_b	n_{iw}/n_i	$(n_{ib}/N_b)*(n_{iw}/n_i)$
1	100	0	10	0.00	0.91	0.00
2	120	10	20	0.07	0.80	0.06
3	40	0	5	0.00	0.89	0.00
4	50	20	50	0.14	0.42	0.06
5	100	20	20	0.14	0.71	0.10
6	30	90	20	0.64	0.21	0.14
						$B_{bw} = 0.36$

- In this case, the probability of a Black person "interacting" with a White person is about 36%. You can also interpret this to mean that 36 of every 100 people a Black person meets will be White.

- Most interaction indexes are not "symmetrical" - the probability of a "typical" Black person meeting a White person is NOT the same as the probability of a "typical" White person meeting a Black one.

- The maximum value of B depends both on the distribution of ethnic groups AND on the proportion of minorities in the city.

- Generally speaking, the value of this index will be highest when the two groups have equal numbers and are spread evenly among tracts.

Measuring Segregation with Multiple Categories: Entropy Index

Both the Index of Dissimilarity and Interaction Index can only measure the segregation of two groups compared to each other. An Entropy Index measures the spatial distribution of multiple groups simultaneously.

The entropy index h for a tract i is:

$$h_i = -\sum_{j=1}^{k} p_{ij} \ln(p_{ij})$$

Where:

k =number of ethnic groups ("ethnicities")
p_{ij} =proportion of population of j^{th} ethnicity in tract i (=n_{ij}/n_i)
n_{ij} =number of population of j^{th} ethnicity in tract i
n_i =total number of population in tract i

Tract	White(1)	Black(2)	Asian(3)	Total pop.	prop. White (p1)	prop. Black (p2)	prop. Asian (p3)	h= - p1*ln(p1)+ p2*ln(p2)…
1	100	0	10	110	0.91	0.00	0.09	0.30
2	120	10	20	150	0.80	0.07	0.13	0.63
3	40	0	5	45	0.89	0.00	0.11	0.35
4	50	20	50	120	0.42	0.17	0.42	1.03
5	100	20	20	140	0.71	0.14	0.14	0.80
6	30	90	20	140	0.21	0.64	0.14	0.89
Total	440	140	125	705	0.62	0.20	0.18	0.92

- The maximum value for h is ln(k), or ln 3 = 1.10
- 0 * log (0) is defined as 0.
- Tracts with higher values of h are more diverse.
  - A tract with h =1.10 would have equal proportions of all groups (33.3 % each).
  - A tract with h = 0 contains only a single group.

The Entropy Index (H) for a City

To compare Entropy Indices between different cities as a whole, White (1986) uses:

$$H = \left(\hat{H} - \overline{H} \right) / \hat{H}$$

$\hat{H}$ = Entropy Index for the city as a whole ("H hat")

	White(1)	Black(2)	Asian(3)	Total pop.	prop. White (p1)	prop. Black (p2)	prop. Asian (p3)	H= - p1*ln(p1)+ p2*ln(p2)…
Total	440	140	125	705	0.62	0.20	0.18	0.92

$\overline{H}$ = the average of the individual tracts' values of h, weighted by population ("H bar"):

Tract i=1,2,3...	h= - p1*ln(p1)+ p2*ln(p2)…	Proportion of total population	Weighted Average
1	0.30	0.16	0.05
2	0.63	0.21	0.13
3	0.35	0.06	0.02
4	1.03	0.17	0.18
5	0.80	0.20	0.16
6	0.89	0.20	0.18
Sum			0.71

$$H = (.92 - .71)/.92 = .226 *$$

(*calculated with nine significant digits, not rounded figures)

- The maximum value of H is 1, when each tract contains only one group ($\overline{H} = 0$). The minimum value of H is 0, when every tract has the same composition as the city ($\overline{H} = \hat{H}$).
- Cities with higher values of H have *less* uniform ethnic distributions. Cities with lower values of H have *more* uniform ethnic distributions.

These descriptions are based on:

Michael J. White, 1986, "Segregation and Diversity Measures in Population Distribution." *Population Index*, Vol. 52, 198-221.

----- 1983. "The Measurement of Spatial Segregation" *American Journal of Sociology* 88 (5): 1008-1018.