

D: Dissimilarity indexes

Frequently used to measure segregation

- How dissimilar are the distributions of two populations?

Example:

4 geographic zones: 1-4

2 types of people: A, B

Group	Population
A:	480 people
B:	160 people

Case 1: completely even distribution

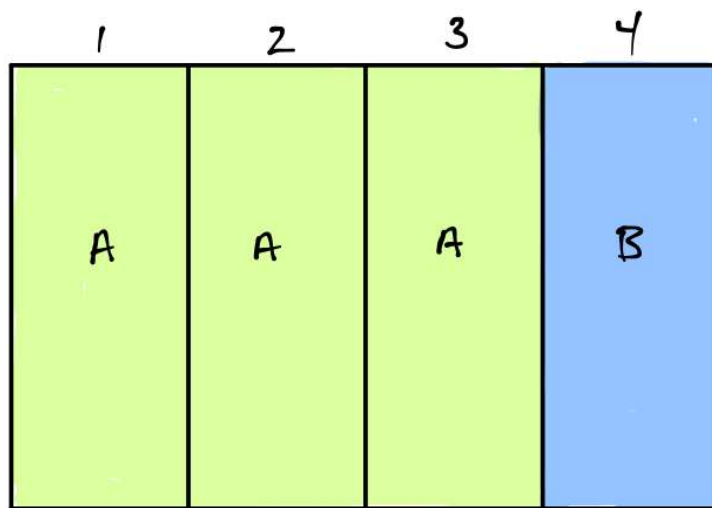
1	2	3	4
A	A	A	A
B	B	B	B

Shares of each population across zones are identical:



Group	1	2	3	4
Share of all A's:	0.25	0.25	0.25	0.25
Share of all B's:	0.25	0.25	0.25	0.25

Case 2: complete segregation



Population shares are very different:

Group	1	2	3	4
Share of all A's:	0.33	0.33	0.33	0
Share of all B's:	0	0	0	1.00

Dissimilarity index:

Summarizes the difference in two distributions

1. Start with the share of each population in each zone

N^t = total number of type- t people

n_i^t = number of type- t people in zone i

s_i^t = type- t people in zone i as a share of all type- t people

$$s_i^t = \frac{n_i^t}{N^t}$$

2. Then sum absolute values of differences over zones:

$$D = \frac{1}{2} \sum_{i=1}^N |s_i^A - s_i^B|$$

Applying to cases:

Case 1: even distribution

Group	1	2	3	4
Share of all A's:	0.25	0.25	0.25	0.25
Share of all B's:	0.25	0.25	0.25	0.25
Absolute difference:	0	0	0	0

$$D = \frac{1}{2} (0 + 0 + 0 + 0) = 0$$

Case 2: complete segregation

Group	1	2	3	4
Share of all A's:	0.33	0.33	0.33	0
Share of all B's:	0	0	0	1.00

Absolute difference:	0.33	0.33	0.33	1.00
----------------------	------	------	------	------

$$D = \frac{1}{2}(0.33 + 0.33 + 0.33 + 1) = 1$$

Features:

- Ranges from 0 to 1 (0% to 100%)
- Larger numbers → greater segregation
- Symmetrical: same result for A-B and B-A
- Independent of relative sizes of the groups, N^A and N^B

Interpretation:

Percent of *one* population that would have to move for equality

Case 2, B moves:

Group	1	2	3	4
Group A	33%	33%	33%	0
B start	0	0	0	100%
B end	33%	33%	33%	0%

Case 2, A moves:

Group	1	2	3	4
A start	33%	33%	33%	0
A end	0	0	0	100%
Group B	0	0	0	100%

Caveats:

- **Any** deviation from uniformity causes $D > 0$
- Random populations (*dissimilar* but no *discrimination*) will have $D > 0$
- D **can be noisy** for small numbers of zones or small populations