Steps for Computing the Dissimilarity, Entropy, Herfindahl-Hirschman and Accessibility (Gravity with Competition) Indices

I. Dissimilarity Index

Measurement: The following formula can be used to measure the evenness between residential use and non-residential land uses.

\[ D = 0.5 \sum_{i=1}^{n} |X_i - Y_i| \]

This implementation of the index, \( D_N \), examines the distribution of land uses at sub-neighbourhood level (i.e., a gridcell) and provides an estimated of its similarity to the distribution pattern at the neighbourhood level. We denote \( n \) as the number of grids in the neighbourhood, \( X_i \) as the ratio of the residential area in grid \( i \) to the residential area in the neighbourhood, and \( Y_i \) as the ratio of the non-residential area in grid \( i \) to the non-residential area in the neighbourhood. The first step to operationalize this index is to divide the neighbourhoods into gridcells with user-defined uniform size which are considered sub-neighbourhoods. Then the comparison of the distribution of land uses within each of the grids to the neighbourhood as a whole can be made. The index of dissimilarity equals to 0 if distribution of land uses within each grid is similar to the distribution at the neighbourhood level, i.e., the proportions of the residential and non-residential land uses within each of the grids are identical to the proportions at the
neighbourhood level respectively. The index equals to 1 if there are either only residential land uses, or there are only non-residential land uses within each of the grids.

NOTE: This measure is different from Cervero’s dissimilarity index measure. See Song and Rodríguez for details.

Steps of Computation:

1. Decide to which neighbourhood each taxlot belongs:
   a. Convert the taxlot polygons into centroids by using XTool;
   b. Spatial join the neighbourhood polygons and the taxlot centroids so that each taxlot would be assigned with a neighbourhood ID and all taxlot centroids that are inside a neighbourhood will have the same neighbourhood ID;

2. Construct grids which would intersect with the study area:
   a. create grids with user-defined size by using “Coordinate Grid Maker 2.290” ArcView script which can be downloaded from ESRI’s website;

3. Assign each taxlot by grid:
   a. Spatial join the grid (polygon) shape file and the taxlot centroids (point) shape file so that all taxlot centroids that are inside a grid cell will have the same grid cell ID;

4. Calculate total area of residential uses for each grid cell by summarizing the table of taxlot centroids shape file based on the grid cell ID. You would obtain a table
with the grid cell ID as the key. Each grid cell also has a neighbourhood ID indicating the neighbourhood to which the grid cell belongs.

5. Calculate total area of non-residential uses for each grid cell by summarizing the table of taxlot centroids shape file based on the grid cell ID. You would obtain a table with the grid cell ID as the key. Each grid cell also has a neighbourhood ID indicating the neighbourhood to which the grid cell belongs.

6. Based on the common grid cell ID, combine the table created from step 4 and the table created from step 5 into a single table. The resulted new table includes the following fields: grid cell ID (key), residential area by each grid cell, non-residential area by each grid cell, and neighbourhood ID.

7. Using the attribute table of taxlot centroids, calculate total residential area by each neighbourhood by summarizing residential area based on neighbourhood ID. This would result a table with neighbourhood ID as the key.

8. Using the attribute table of taxlot centroids, calculate total non-residential area by each neighbourhood by summarizing non-residential area based on neighbourhood ID. This would result a table with neighbourhood ID as the key.

9. Based on the common grid cell ID, combine the table created from step 7 and the table created from step 8 into a single table. The resulted new table contains the following fields: neighbourhood ID (key), residential area by each neighbourhood, and non-residential area by each neighbourhood.

10. Based on the common neighbourhood ID, join the table created from step 9 to the table created from step 6. In addition to the existing fields in the table created
from step 6, the new table will also have information on total area of neighbourhood residential and non-residential uses by each grid cell.

11. In the table created from step 10, divide total residential area in grid cell by total residential area in neighbourhood. A new field containing the information on the ratio is added to the table.

12. In the table created from step 11, divide total non-residential area in grid cell by total non-residential area in neighbourhood. A new field containing the information on the ratio is added to the table.

13. Calculate the difference between the two ratios obtained from step 11 and 12 and add the new field - ratio difference - to the table.

14. Calculate Dissimilarity Index:

   a. Use the table created from step 13 to summarize the grid cell ratio differences by neighbourhood ID;
   
   b. Divide the summarized total differences by 2 for each neighbourhood. This is the value of dissimilarity index for each neighbourhood.

II. Entropy measures

The entropy index (Shannon Index) is commonly calculated through the following formula:

\[
Entropy = \left(-\sum_k \left[ (p_1) \ln(p_1) \right] \right) / (\ln k)
\]
Researchers in various fields have implemented this measure in multiple ways to examine the dispersion of objects at interests. Examples include:

*Measurement:* To examine the distribution pattern of different land uses within a neighbourhood, the index spells out that $p_i$=proportions of each of the six land use types such as single family residential, multi-family residential, commercial, industrial, public institutional and park uses, and $s=$the number of land uses. In this case, $s=6$ (Song and Knaap, 2004).

**Steps of Computation:**

1. Decide to which neighbourhood each taxlot belongs:
   a. Convert the taxlot polygons into centroids by using XTool;
   b. Spatial join the neighbourhood polygons and the taxlot centroids so that each taxlot would be assigned with a neighbourhood ID and all taxlot centroids that are inside a neighbourhood will have the same neighbourhood ID;

2. Calculate the total area of each of the six land use types in each neighbourhood:
   a. Using the taxlot centroids attribute table, summarize the area of taxlots by each land use type and by neighbourhood ID. The resulted table includes the following fields: neighbourhood ID, land use type, and the total area by land use type.
3. Calculate total area of all land uses in each neighbourhood:
   a. Using the taxlot centroids attribute table, summarize the area of taxlot by
      neighbourhood ID.

4. Based on the common neighbourhood ID, join the table created from step 3 to the
   table created from step 2. The new table includes the following fields:
   Neighbourhood ID, land use type, area by land use type, and total neighbourhood
   area.

5. Using the table created from step 4, divide the area of each single land use type by
   total neighbourhood area. New fields – proportions of each land use type \( \frac{P_i}{\sum} \) –
   are added to the table.

6. Calculate \((p_i)(\ln p_i)\) and create the new fields with the values.

7. Calculate \(\sum_i [(p_i)(\ln p_i)]\) : Summarize the field \((p_i)(\ln p_i)\) for all 6 land use
   types by neighbourhood ID and create new fields.

8. Finally calculate entropy index for each neighbourhood using the formula:
   \[ Entropy = \frac{-\sum_i [(p_i)(\ln p_i)]}{\ln k}. \]

III. Herfindahl-Hirschman Index (HHI)

Following the same notation as the entropy index, the formula of calculating HHI among
\( k \) uses of land is:

\[ HHI(k) = \sum_{i=1}^{K} (P_i \times 100)^2 \]
where \( p_i \) is the percentage of each type of land use in the neighbourhood, and \( K \) is the number of land use types.

**Steps of Computation:**

1. Using the table created from step 4 in the entropy index, divide the area of each single land use type by total neighbourhood area. New fields – proportions of each land use type \( (p_i) \) – are added to the table.

2. For each newly created field in step 1, calculate \( (p_i \cdot 100)^2 \) and create \( i \) new fields with these new values.

3. Calculate \( \sum_{j=1}^{K} (P_i \cdot 100)^2 \): Summarize the field created in step 2, \( (P_i \cdot 100)^2 \), for all \( k \) land use types by neighbourhood ID and create new fields.

IV. **Gravity-based Measure with Competition**

**Measurement:**

\[
AG_i = \sum_{j=1}^{n} A_{ij} d_{ij}^{-\beta} / \sum_{k=1}^{m} d_{kj}^{-\beta}
\]

where
$AG_i =$ accessibility of residential land use (e.g., house units or neighbourhoods) $i$ to non-residential land use (e.g., non-residential land parcels or activity centres) $j$

$A_j =$ attractiveness of non-residential land use $j$, in this case, it’s measured by floorspace of retailing stores

$d_g$ or $d_{ki} =$ distance from residential use $i$, or housing unit $k$, to non-residential use $j$

$d_{ij}^{-\beta}$ or $d_{kj}^{-\beta} =$ impedance function based on the inverse power function

$\beta =$ distance decay parameter

$n =$ number of one type of non-residential land uses

$m =$ number of residential land uses

For example to calculate single-family household’s accessibility to retailing, there are $n$ commercial retail stores and $m$ houses in the study area. $AGi$ is the accessibility level of household $i$ to retail services. Service supply is measured by the floor space of retail stores. In other words, if a certain store provides more floor space, it is assumed that store has a larger capacity for goods and customers, and therefore provides a higher level of accessibility to area residents. This supply capacity is weighted by the distance that an individual customer has to travel. If the store is farther from home, household members are less willing to take a trip to that store, and therefore that establishment provides a lower level of accessibility.
Steps of Computation:

1. Import the attribute table of single family taxlots to Microsoft Access database named as “db1.mdb” and name the table as “SFR”.

2. Import the attribute table of commercial stores to “db1.mdb” database and name the table as “STORE”.

3. Create a new table named as “RES” to store the information on distances and store floorspaces. Modify the table by applying the developed C++ program.

4. Create a new table named as “NEWSTORE” to store the information on the summation of distances among single family housing units to commercial stores. Apply the developed C++ program to obtain the values.

5. Create a new table named as “ACCESS” to store the final values of the index. Apply the developed C++ program to obtain the final values.

Note: You need Microsoft Visual C++ along with ADO to run the developed C++ program. The code snippet is available upon request.