

# SDGB/DGGB 7844 HW 2: Townsend Material Deprivation Index

Instructor: Prof. Nagaraja

Due: 10/15 in class

Submit your homework as follows: (a) hardcopy of assignment only in class; (b) e-mail assignment and code to [fordhamRcomputing@gmail.com](mailto:fordhamRcomputing@gmail.com) by the start of class. In your e-mail, use the subject heading: “HW 2: [Your name]” and include HW 2 and your name the file names you attach. Comment your code for full credit. Complete your work individually.

Deprivation indices are used to measure levels of deprivation in populations which go beyond simply measuring income (as in poverty measures). Initially developed in the UK, the most common are the Townsend, Carstairs, Jarman, and the Index of Multiple Deprivation. These indices have been shown to be good proxies for indicating health-related problems in local communities<sup>1</sup>.

Goal: In this assignment, you will compute the Townsend Material Deprivation Index for census tracts in New York County (i.e., Manhattan) using American Community Survey (ACS) data (administered by the U.S. Census Bureau).

The Townsend index is constructed from four variables: % unemployment among people 16 and over (unemp), % of overcrowded households defined as homes with more than one person per room (oc), % of housing which is rented (rent), and % of households with a vehicle (car). A higher percentage of any of these variables indicates more deprivation in a geographic area (Note: you should not convert the percentage variables into decimal form). Let geographic region be denoted by the subscript  $i$  and variable by  $j$ . To compute the index, follow the four steps described next.

---

<sup>1</sup>For more information: (a) Townsend, P., Phillimore, P., and Beattie, A. (1988). *Health and deprivation: inequalities and the north*. Croom Helm, London. (b) Carstairs, V. and Morris, R. (1991). *Deprivation and health in Scotland*. Aberdeen University Press, Aberdeen. (c) Jarman, B. (1983). Identification of underprivileged areas. *BMJ* **286** 1705-1709. (d) Jordon, H., Roderick, P., and Martin, D. (2004). The index of multiple deprivation 2000 and accessibility effects on health. *Journal of Epidemiology and Community Health*. **58** 250-257.

(a) transform the variables to reduce skewness (note: log is natural log):

$$\begin{aligned}t_{i,unemp} &= \log(x_{i,unemp} + 1) \\t_{i,oc} &= \log(x_{i,oc} + 1) \\t_{i,rent} &= \log(x_{i,rent} + 1) \\t_{i,car} &= \sqrt{x_{i,car}}\end{aligned}$$

(b) compute geographic region mean  $\bar{t}_j$  and standard deviation  $s_j$  for each variable  $j$ .

(c) standardize variables for each region  $i$  and variable  $j$ :  $z_{ij} = \frac{t_{ij} - \bar{t}_j}{s_j}$

(d) compute sum of standardized variables for each region  $i$ :  $Townsend_i = \sum_{j=1}^4 z_{ij}$

You will have one Townsend index value for each region  $i$ . These values are on a relative scale; that is, the actual number is meaningless, just the ranking is important. A negative score indicates a less deprived region; a score of 0 indicates roughly the average level of deprivation; a positive score indicates a more deprived region.

1. What is a census tract? How many census tracts are in New York County?
2. Describe one advantage and one disadvantage of computing estimates after combining 5-years of data.
3. Download the ACS data for the 2008-2012 5-year estimates for all New York county census tracts for the following variables using American FactFinder ([link](#)). Table DP03 contains selected economic characteristics and Table DP04 includes selected housing characteristics.
  - unemployment: Table DP03, variable HC03\_VC08
  - housing tenure (whether house is rented or owned): Table DP04, variable HC03\_VC64
  - no vehicles: Table DP04, variable HC03\_VC82
  - low occupancy: Table DP04, variable HC03\_VC110 (note: you will have to transform this variable to get % overcrowded to use in the index)

Merge the tables into one data frame, each row representing a census tract, each column representing one of the Townsend variables.

For each variable, construct a histogram and compute the following summary statistics: mean, median, standard deviation, maximum, and minimum. Describe the shape of each histogram.

4. How many observations are missing for each variable? What percentage of census tracts do not have complete data? Is this a problem for our analysis? Justify your answer. (Note: do not delete tracts with missing data.)
5. Construct a scatterplot matrix of the four variables. Are they linearly related? Now, transform the variables as given in step (a), adding the transformed variables to your data frame. Make another scatter plot matrix with the transformed variables. Are they linearly related? Construct a correlation matrix of the transformed variables and describe your results. Are high or low correlations preferable when constructing an index? Justify your answer.
6. Compute the Townsend index value for each census tract. Add the index to your data frame. Which census tract is the most deprived and which is the least deprived (give the census tract number and deprivation index level)?
7. The ACS data includes not only estimates but their margins of error which we ignored in our calculations. What are the implications?
8. Construct a map color-coded by the deprivation index value. Download the shape files for New York state census tracts for 2012 from the U.S. Census Bureau website ([link](#)). Extract the tracts for New York County only. In your map, go from red for the most deprived areas to blue for the least deprived. Include a legend and plot title. Describe the patterns you see, especially in relation to what you know about neighborhoods in New York City. What does the large rectangle in the middle of the map represent?
9. In which census tract is Lowenstein (where we have class)? What is the deprivation level rank (where a rank of 1 is the most deprived)? Mark it on the map (use the computer, not by hand) and add it to your legend.
10. New York County is an urban county, however New York state has roughly 22 counties classified as rural (e.g., Allegany, Essex, Otsego, Sullivan). Would it make sense to compute the Townsend index values for all census tracts within New York state combined? Why or why not?