1 Market-level data, neighborhood choice

This part of the assignment is based on "White Flight and Coming to the Nuisance: Can Residential Mobility Explain Environmental Injustice " by Depro et al. (2015).

To better understand the neighborhood dynamics underlying the observed changes in aggregate demographics following a change in enviromental risk/exposure, Depro et al. (2015) use a more structural model of neighborhood choice. They aim to identify the role of NATA total cancer risk in driving residential mobility. The probability that a member of a particular group in tract k will choose to move to tract j, $P_{j,k}$, is defined as a function of location attributes.

There are a lot of moving parts in this paper. This assignment focuses on estimating the δ and μ parameters. Note that this estimation is carried out separately for each demographic group.

1.1 Comprehension

1. **Question:** Describe in words what the δ parameters represent. Provide a verbal explanation of how these δs are estimated using Equations 10, 11, and 12.

1.2 Data construction

There is a set of code fragments in the assignment folder (document titled code fragments). This code creates a data frame in R that can be used to calculate the delta vector for white residents. See **Code Fragment 1**. Data file can be found in the same folder.

- Running the code takes a long time if all 2000 census tracts are included. To make this easier, code drops all tracts except for the first 20.
- The variable mc (moving cost) is created for each tract, but in equation 11, moving cost is indexed by the tract you move from (k) to the tract you move to (j).
- The rbind command adds a single row to the table which is the outside option (which in the full model with all census tracts would represent the rest of the country).
- To make the delta vector for Hispanic residents, change nwnh00 and nwnh10 to nh00 and nh10.

Questions:

1. The final line of **Code Fragment 1** which creates the sigmas in equation 11 is incomplete. Complete this last step of the code (and run it) on your own.
2. What are the only variables remaining in the cleaned data frame? How are these implicated in the estimation of the vector of deltas?

1.3 Contraction mapping

The code includes a function that starts with a guess for the δ vector and μ and returns a vector of that match population shares by neighborhood in 2010 conditional on the μ guess (see **Code Fragment 2**).

- The function uses map commands from the r package purrr, which you may be unfamiliar with, in order to accomplish some of the more complex summations in equation 11.

- Notice the function includes a counter, which allows you to see how many times the function must re-guess the delta vector, and a kill number, which will stop the function after a prescribed number of loops.
- If you provide a particularly bad guess of μ, the function may never converge, so the kill input can be useful.
- Notice that in equation 11, l=k and j=k are special cases.

Questions:

1. Five parts of Code Fragment 2 are left blank. Fill them in (and briefly explain what these missing pieces do).

   HINT: The $MC_{l,k}$ used to calculate the denom vector should be: '$(.y + dfmc[[l]] - 11.85046)$'

2. Take the following steps:

   - Set kill=1. Notice, in the first run, δ_0 is randomly generated. What do the predicted population shares like? How different are the new delta guesses ($delta_1$)? Why do some tracts have higher values of δ_1?
   - Now run the code with kill=2 and kill=3. How accurate are the predicted population shares now? How much difference is there between the $delta_0$ and $delta_1$?
   - Now run the function with kill=1000. Hopefully, the equation should converge after about 20 to 25 loops. What exactly is happening when it 'converges' (ie: what does it mean that sum(dfδ_{fail})==0).
   - Now change the value of mu_{guess} to .001 and .01. How does this change the population shares? Explain why this happens.