Exercise 5

Confounding factors & variable adjustment

In statistics, a confounding variable is an extraneous variable in a statistical model that correlates with the dependent variable and the independent variable. This type of relationship is sometimes termed a spurious relationship. When we evaluate the importance and the nature of a risk to human health, it is key to control for confounding variables in order to isolate the effect of a particular factor such as a sugar sweetened beverages, pesticide, or income.

To adjust a variable of interest and remove the effect of an associated confounding factor, subtract the predicted value from the raw value of the variable of interest and then add the median of the variable of interest for the investigated population.

Table - bmi.income.5435.lsne

	id	x	y	revmed	bmi
1	1	536671.940000	153764.970000	51197.000000	26.397266
2	2	537731.500000	151423.970000	54265.000000	27.699286
3	3	537317.940000	153974.090000	58926.000000	26.245810
4	4	538103.880000	151514.730000	60710.000000	24.050879
5	5	538692.750000	151722.360000	73121.000000	30.482939
6	6	537651.000000	151784.000000	54265.000000	28.117283
7	7	539529.000000	151674.980000	45281.000000	21.461937
8	8	539009.130000	151890.130000	73121.000000	27.580200
9	9	538339.940000	152191.220000	65883.000000	30.117630
10	10	537991.940000	151797.920000	55686.000000	24.459284
11	11	538502.940000	153873.250000	57252.000000	23.566631
12	12	537433.940000	152860.950000	45138.000000	24.948097
13	13	536629.310000	153895.160000	51197.000000	26.988636
14	14	539431.940000	151854.050000	73121.000000	28.408163
15	15	539311.000000	152320.860000	73575.000000	22.500000
16	16	539628.940000	152963.110000	50882.000000	26.407543
17	17	537004.940000	154163.140000	53519.000000	25.709877
18	18	537778.440000	151764.000000	54265.000000	26.467010

Tasks

The zip file “exercise5.zip” is available on Moodle. It contains a shapefile named “bmi.income.5435.lsne.shp”. This file has 2 attributes: the median income (revmed) and a raw BMI value (bmi) (see table above).

1. Produce a scatterplot of the relationship between BMI (y) and the median income (x); copy the scatterplot and paste it in a ms-word document (for your short report);

2. Transform the BMI variable so that it becomes adjusted for the median income. Apply a linear regression and use the median of BMI for the whole population. You can calculate the regression with the software you want (R, Matlab, Excel)
3. Then import the adjusted BMI in Geoda so that you have an additional column named "adj.bmi";
4. Produce a weight file corresponding to a spatial lag of 400 meters;
5. Process the Getis-Ord Gi* statistics (weight row-standardized) for the raw BMI variable (you can add a basemap to benefit from a geographic background);
6. Copy and paste the map in your short report;
7. Process the Getis-Ord Gi* statistics (weight row-standardized) for the adjusted BMI variable (you can add a basemap to benefit from a geographic background);
8. Copy and paste the map in your short report;
9. In your short report, describe where are the raw and adjusted BMI hotspots and explain the main differences in the spatial patterns of these two variables ;