Sean Connelly
UPP 465
Professor Tilahun
4/6/20

Spatial Statistics – Autocorrelation Assignment

1. Researchers sought to examine if the pattern of industries in North Eastern United States formed spatial knowledge clusters. A measure they thought would capture knowledge creation at the firm level was the number of patents granted over the prior 3 years. Their analysis planned to use county level data for 60 counties (N=60). They start their analysis by first assessing global autocorrelation. Based on this information, answer the following questions:
    a. What variable should they use for their spatial analysis at the county level? Fully explain your reasoning.

       The researchers should use the average number of patents per firm at the county level to conduct their spatial analysis. By normalizing for the number of firms and making the variable of interest a rate, any pattern will instead hopefully reveal if there is a spatial association behind knowledge clusters, rather than just reflecting the overall population or number of firms.

    b. Briefly explain what type of neighbor definitions you would use? Why?

       Among the traditional boundary definitions, a Queen criteria is the only logical option. Knowledge spillovers do not abide by the strict limitations of Rook or Bishop – it can be spread anywhere. A K Nearest Neighbors definition might leave out logical options solely due to the areal shape of the county and is not ideal. A distance-based metric built on underlying theoretical research on knowledge diffusion would be the most appropriate.

    c. Suppose the researchers used two different weight matrices: one based on a nearest k neighbors (k=2) definition and another using an inverse distance weight matrix to compute the Global Moran's I. They found the following results:

Type	I	N	E(I)	Z-stat	p(>Z)
Nearest K	0.3	60	-0.0169	2.5	0.006
Inv. Distance	0.19	60	-0.0169	3.1	0.001

- Compute what the expected value of the global Moran's I E(I) is under the null hypothesis.

  $E(I) = -1 / (n - 1) = -1 / (60 - 1) = -1 / 59 = -0.01694915254$

- Interpret both results. Which result shows stronger evidence of spatial autocorrelation?

  Both values of Global Moran's I are positive, suggesting some evidence of positive spatial autocorrelation, or that similar values are observed between neighbors. However, the result using an inverted distance weight matrix has a more extreme *Z statistic* and a lower *p value*, indicating a lower likelihood that the such a pattern would be seen under spatial randomness to that level of statistical significance.

2. Suppose you were undertaking an analysis of local spatial autocorrelation for a variable of interest X. The Moran scatter plot is as shown in the figure below where the dotted lines indicate the means of the variable and its spatial lag. Answer the following questions (see next page for plots):
   a. Describe what the spatial lag variable indicates.

      The spatial lag variable indicates the average of the variable of interest (in this case, X) in neighboring locations, defined by whatever criteria is implemented.

   b. What type of spatial autocorrelation does the figure suggest? Random, positive, or negative?

      The figure indicates positive spatial autocorrelation.

   c. In figure 1, label the areas of positive and negative spatial autocorrelation.

      See below.

   d. In figure 2, label the four quadrants in terms of the relationship of the variable with its spatial lag. Note figure 1 and 2 are the same.

      See below.

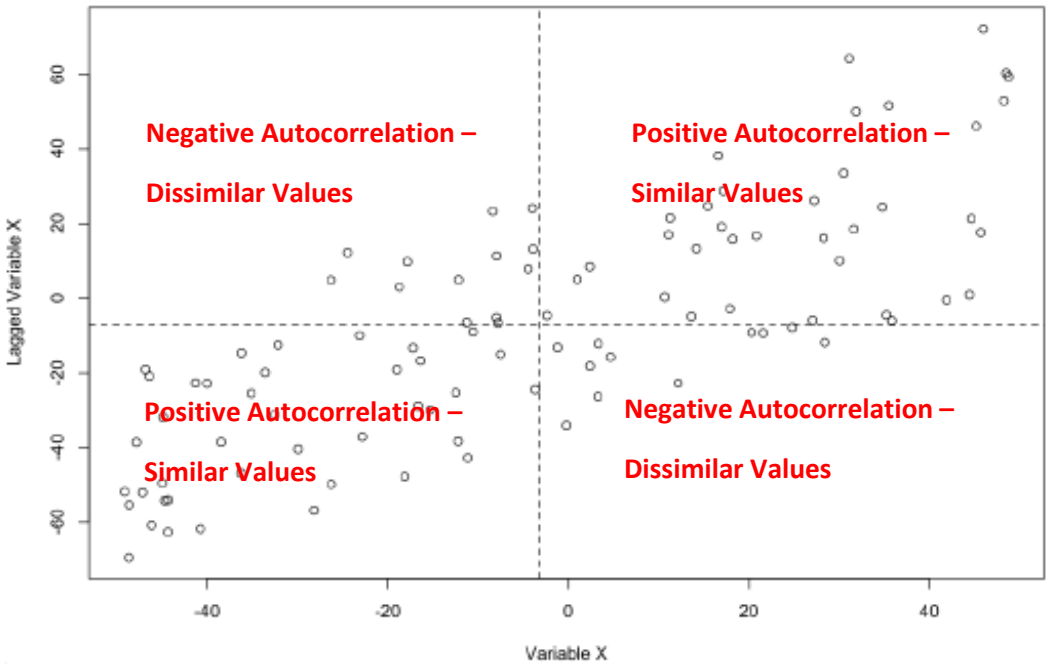

Figure 1: Moran's Scatter Plot for problem 2C

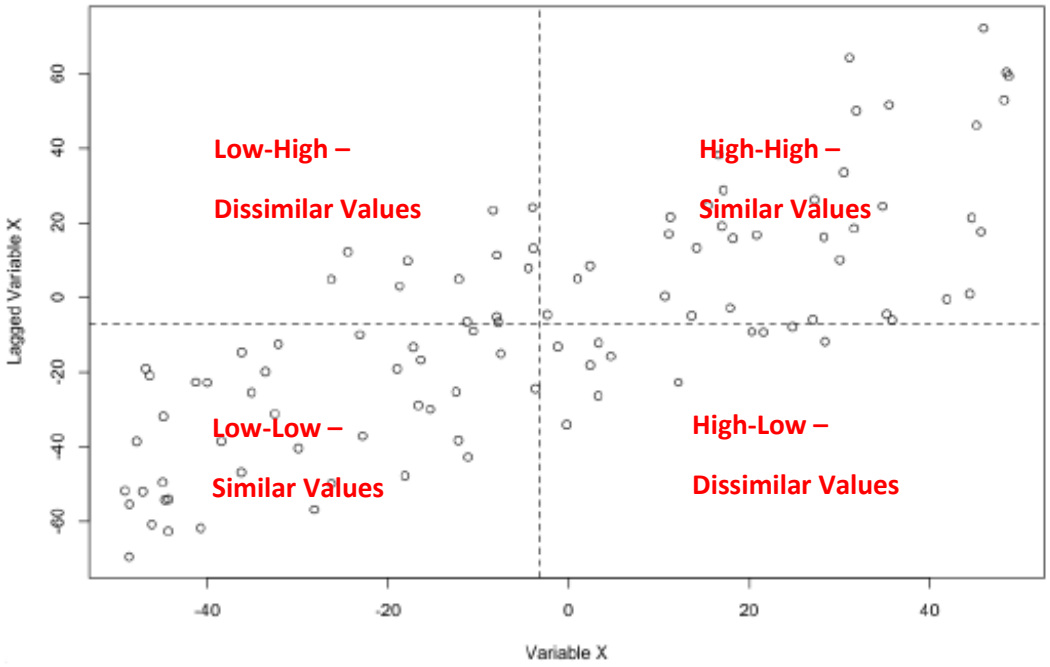

Figure 2: Moran's Scatter Plot for problem 2D