**GGRC32 – Essential Spatial Analysis**

**Assignment 3: Area Pattern Analysis and Spatial Autocorrelation**

**Fall Semester 2017**

By: Grisham Nathan

**Due November 7th before class**

**Submit via Blackboard**

**Part I: The Join-Count Statistic**

1. Researchers want to know if high/low voter-registration rates are randomly distributed. Given the figure below, answer the following questions using *rooks case* and the *nonfree sampling test procedure*. Black indicates a high value, white a low one. (8 pts. total)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* 1. Fill out the following table, counting the number of joins of each type: (1 pt.)

|  |  |
| --- | --- |
| **W/W** | 3 |
| **B/B** | 3 |
| **B/W** | 18 |
| **Total** | 24 |

* 1. What are the null and alternative hypotheses? (1 pt.)

H0: OBW = EBW; The pattern is random

HA: OBW ≠ EBW; The pattern is not random

* 1. What is the expected number of B/W joins for this distribution? (1 pt.)

EBW = 12.8

* 1. What is the standard error of the expected B/W joins (σBW)? (2 pts.)  
     X = 416/15

Y = 128.6564103

σBW =2.312951282

* 1. What is the Z-score? (1 pt.)

Z = 2.24820991279

* 1. At α = 0.05, what is (or are) the critical Z-scores? (1 pt.)

The two sided Join count hypothesis is what I am conducting. The critical Z-scores are -1.96 and 1.96

* 1. What is your decision regarding the null hypothesis? (1 pt.)

The Z-score is not in between the two critical Z-scores. As a result, we reject the null hypothesis that the pattern is random at 5% significance level.

**Part II: Moran’s I**

1. **Interpretation of Moran’s I values** (3 pts.)
   1. What does a value of 1 indicate about spatial autocorrelation and spatial patterns?

It indicates a strong level of positive spatial autocorrelation. This means that high values are located are close to high values and low values are located to low values. The spatial pattern is clustered.

* 1. What does a value of -1 indicate?

It indicates a Strong level of negative spatial autocorrelation. This means that high values are located close to low values. The spatial pattern is uniform.

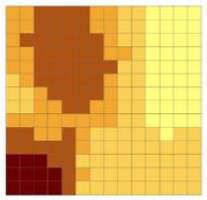
* 1. What does a value of 0 indicate?

It indicates no autocorrelation which means that the spatial pattern is random. High values and low values are randomly distributed on the map.

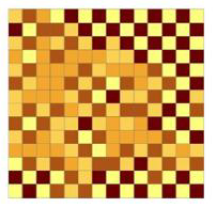
1. Look at the following maps (dark colors represent high values, and light colors represent low values). Answer the associated questions with each map. (3 pts.)
   1. What will the Moran’s I value be closest to?
      1. 0
      2. +1
      3. -1

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* 1. The map below is likely to have what type of Moran’s I value?
     1. Insignificant and negative
     2. Insignificant and positive
     3. Significant and negative
     4. Significant and positive



* 1. The map below is likely to have what type of Moran’s I value?
     1. Significant and positive
     2. Insignificant and negative
     3. Significant and negative
     4. Insignificant and positive



1. **Z-test with Moran’s I**: Researchers are looking at the spatial autocorrelation of the percentage of black citizens in each census tract in Milwaukee. There are *412* tracts in Milwaukee that they measure. They find that **Moran’s I = 0.820, and Var(I) = 0.13**. (4 pts. total)  
   \*\*\* To answer part (a), you will need to use the equations:

* 1. Calculate the Ztest statistic of this dataset. (2 pts.)

I – E(I) = 0.82+(1/411)

Ztest = (0.82+(1/411))/sqrt(0.13)

Ztest =2.281018982

* 1. If we are conducting a one-tailed test where the alternative hypothesis is that there is positive spatial autocorrelation of the data, what is our *critical* Z-score at α = 0.05? (1 pt.)

Zcrit = 1.645

* 1. At α = 0.05 with a one-tailed test, what is our decision regarding the null hypothesis given our Ztest statistic? (1 pt.)

Zcrit < Ztest

We reject the null hypothesis that the data has no autocorrelation or Moran’s I = 0 at 5% significance level.

1. Compute Moran’s I using rook connectivity for the following map. What type of spatial autocorrelation, if any, exists in this study area? Show your W matrix and all calculations. (10 pts.)



**Shown in excel spreadsheet**

**Part III: Neighbourhood Weight Matrix Analysis (in ArcMap)**

For this part of the assignment, we will be creating and investigating a weights matrix. Begin by finding the *Generate Spatial Weights Matrix* tool in the Spatial Statistics Toolbox. Use this tool to save a spatial weight matrix (SWM) file for the Toronto wards to your USB drive. Choose *wardnum* as the unique ID field. Use the conceptualization corresponding to 1st order queen contiguity. You can use common sense to figure out which option to choose. Or look back to the lecture notes. Make sure you choose to create a row-standardized weight file. You may leave all other options blank and click OK.

The SWM file can be used as an input item in many of the spatial statistics tools. This can be very helpful as creating this matrix can take a lot of time on larger datasets. For now, we need to apply the *Convert Spatial Weights Matrix to Table* tool so that we can visualize the neighbourhood sizes. Run the tool, saving the new table to the new folder. Next, add this new table (a .dbf) file as a layer to the map. View the attribute table of this new table and answer question 1 below. Right click on the WardNum column and Summarize the WardNum field. You don’t need to select anything special in the summarize options, but make sure to save the file to your folder and add the table to the map. Answer question 2.

Finally, join the summarized table to the *TorontoWards* table. This is accomplished by right-clicking on *TorontoWards* and choosing Joins. You can join the two tables on the *WardNum* field. Answer questions 3 & 4.

Questions for Part III

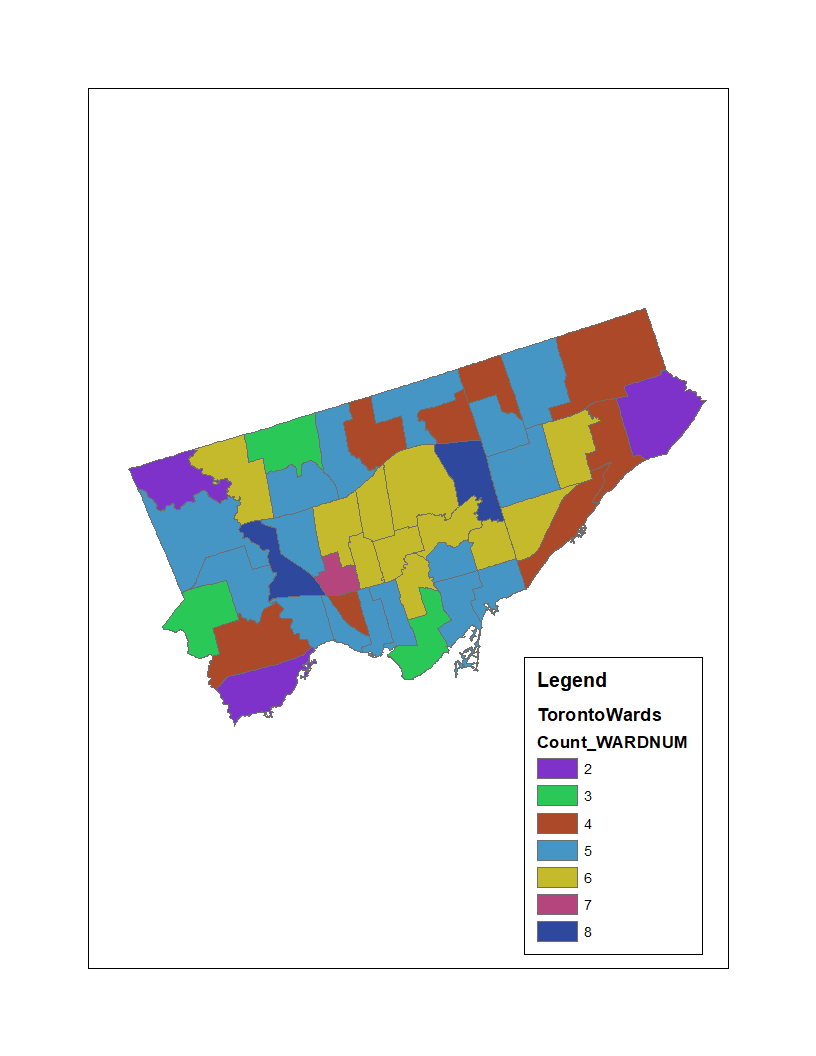
1. Sort the spatial weights table by the WardNum column. Describe what the WardNum and NID columns are keeping track of. (1 pt.).

WardNum is the column that shows the ward number for each ward. NID is the column that shows the ward numbers for the wards that connect with the corresponding ward of each WardNum.

1. Sort the summarized table by *WardNum*. What is being measured in the *Count\_WardNum* field. (1 pt.).

In the Count\_WardNum field, the amount of wards/NIDs that connect to each WardNum is displayed.

1. Create a choropleth map of the Toronto wards indicating the number of nearest neighbours each ward has. You may do this using the categories|unique values symbolization. (2 pts.)



1. Describe the pattern you see of high and low neighbour counts (1 pt.). Which spatial “pitfall” is causing this pattern (1 pt.).

The wards with 4 neighbours, 5 neighbours and 6 neighbours tend to be more clustered than the wards with other neighbour counts. The amount of wards that have below 4 neighbours or more than 6 neighbours tend to be very low and are more spread out on the map compared to the wards with neighbour count between 4 and 6. Wards with neighbour counts between 4 and 6 tend to be very clustered and are high in amount. The spatial “pitfall” causing this pattern is spatial autocorrelation.

**Part IV: Moran’s I (in ArcMap)**

Start by opening the TorontoWards table and add a new field stored as a double. Call it PCT\_FORD. Now calculate the percentage of votes that went to Ford ( ie: 100 \* Ford / (Ford + Smitherman + Pantalone) ). Repeat this process, calculating the percent of votes that went for each of the three candidates. After you’re done, look at the first row in the table: if the calculations were done properly, the three percent fields should add to 100. Answer Question 1 below.

Right click on TorontoWards and remove all the joins. Next you will calculate Moran’s I for each of the percentage variables just created, using the queen weight matrix (.swm file) we created in part 1. Use the *Spatial Autocorrelation* tool to compute Moran’s I for the percentage of Ford voters in the TorontoWards feature class. For the *Conceptualization of Spatial Relationships*, choose the option to *Get Spatial Weights from File* and browse for the weights matrix file. Do not run the option for graphical output (it can be buggy). Record the Moran’s I value along with the Z-score and p-value. Repeat the process for the variables you created for the two other candidates.

Answer the remaining questions below.

Questions for Part IV

1. Which Ward had the lowest percentage of voters for Ford? Which had the highest? (2 pts.)

**Lowest**

Ward Name: Trinity-Spadina

WardNum: 19

**Highest**

Ward Name: Etobicoke North (2)

WardNum: 2

For Pantalone? (2 pts.)

**Lowest**

Ward Name: Etobicoke North (2)

WardNum: 2

**Highest**

Ward Name: Trinity-Spadina

WardNum: 19

And for Smitherman? (2 pts.)

**Lowest**

Ward Name: Etobicoke North

WardNum: 2

**Highest**

Ward Name: Toronto Centre-Rosedale

WardNum: 27

1. Record the Moran’s I, Z-score, and p-value for the three variables. (3 pts.)

**Ford**

Moran’s I = 0.699720

Z-Score = 7.670137

P-value = 0

**Smitherman**

Moran’s I = 0.652037

Z-Score = 7.30591

P-value = 0

**Pantalone**

Moran’s I = 0.610897

Z-Score = 6.89767

P-Value = 0

1. Interpret the results. What can the values of Moran’s I tell us about the spatial patterns of voting? (3 pts.)

The P-value for each variable is 0, which is lower than 0.05 significance level. Therefore, we can say that there is strong evidence that the Moran’s I doesn’t equal 0 (0 means no autocorrelation) and the spatial patterns of the data are not random. In addition to the 0 p-values, Z-Scores are all positive, which means that there is strong evidence that spatial voting patterns are more clustered than random. The values of Moran’s I are between around 0.6 and 0.7, which indicates that the data has a moderate level of positive spatial autocorrelation. As a result, we can conclude that the spatial patterns of voting is more clustered than random.