Assignment I - Spatial Autocorrelation

DA3480 - Spatial Data Analytics

Department of Decision Sciences
University of Moratuwa

Assignment Briefing

You need to submit an analysis report containing 500-1000 words and appropriate figures (or snapshots) to answer the following general questions and the specific questions raised in the steps. The analysis report should follow the following outline.

- Title : Exploring Spatial Autocorrelation and Moran's Index Analysis of <hotel review scores>
- Author Name and Index No.
- Introduction : Refer to the "HotelLocation_ColomboGT300Reviews.csv" file and describe the characteristics of the dataset.
- Objectives :
  - ▶ Determining if there is evidence of spatial autocorrelation or dispersion in the review scores of the hotels.
  - ▶ Calculating Moran's index as a measure of spatial autocorrelation.
  - ▶ Providing a data-driven recommendation for the stake holders on the spatial dependency of the review scores.

Assignment Briefing

1. Before Analysis Setup :
   - What is the purpose of setting a seed for randomization ?
2. Creating Distance-Band Weights :
   - How do you reproject the provided dataset to UTM projection ?
   - What are the steps for creating a weights matrix in GeoDa ?
3. Inference :
   - How do you perform randomization with a specified number of permutations ?
   - What is the purpose of setting the number of permutations ?
   - What does the reference distribution for the statistic depict, and how is it represented ?
   - How can you use the distribution to draw conclusions about the analysis ?
4. Moran Scatter Plot Options :
   - What options can be accessed by right-clicking on the Moran scatter plot ?
   - What steps are involved in computing a standardized version of the Review Score and its spatial lag ?

Steps to Follow

Random Seed

Follow these steps before you start the analysis :
- Open GeoDa and go to **File** –> **Preferences** menu.
- In the menu, locate "Set Seed for Randomization"
  - ▶ Click on "Specify Seed..." box in the options menu.
  - ▶ A dialog box will appear, displaying the default seed value (usually 123456789).
  - ▶ Enter your index number without the English letter as the specified seed value (e.g. the holder of the index number 123456A should enter 123456 only)
 - ▶ Specify the number of CPU cores you want to use (e.g., 6) in the provided field.
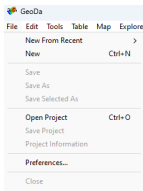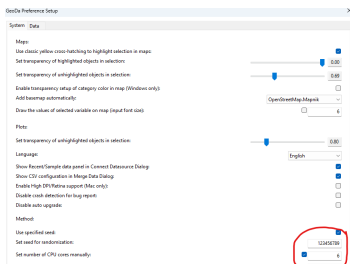
Creating distance-band weights

- Use "HotelLocation_ColomboGT300Reviews.csv" and QGIS to generate a point data set.
- Use QGIS to reproject the provided dataset to UTM projection. Save the reprojected file as "indexno_hotellocations.shp"
- Import the reprojected ".shp" file to GeoDa
- Create the weights matrix following the steps below
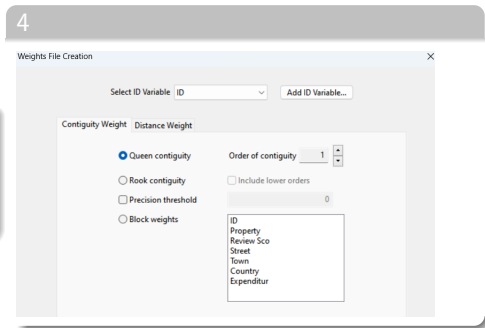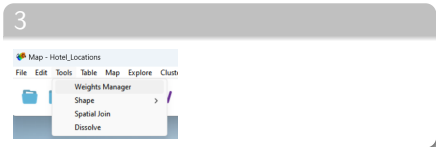
Creating distance-band weights

5

Weights File Creation

Select ID Variable ID ⌄   Add ID Variable...

Contiguity Weight | Distance Weight

Variables:
Geometric centroids | Variables
  X-coordinate variable Long ⌄
  Y-coordinate variable Lat ⌄

  Distance metric: Euclidean Distance ⌄

Method:
Distance band | K-Nearest neighbors | Kernel

  Specify bandwidth   0.037219

  ☐ Use inverse distance?   Power   1 ⇕

6

Click on the **Create** button and specify a file name, indexno_weights (the GWT file extension is added automatically), the new weights and their summary properties are listed in the **Weights Manager**
Visit **Tools** −> **Weights Manager** and take a print-screen of the table (you need to include this on your analysis report).

Distance Metric

Answer the following questions regarding the Distance Metric

- Why do you choose Euclidean distance when creating the Distance Metric ?
- What is the critical distance for your point data ?
- What is the meaning of "critical distance" ?
- Explain the importance of the "threshold value" ?

Create the Moran's index using the Review Score as the spatial variable.

Interpretation

Open the Moran's index scatter plot. The plot shows a linear fit through the point cloud. The slope of this line corresponds to Moran's I. Answer the following questions.

- What can you tell about the shape of the point cloud ?
- Identify locations in the map (or, in any other open view) associated with each of the four types of spatial autocorrelation. What interpretations can you make using the available information ?

Inference

Right click on the scatter plot and select **Randomization** and 999 number of permutations in the side panel. The permutation value corresponds to the M in the expression for the pseudo p-value.
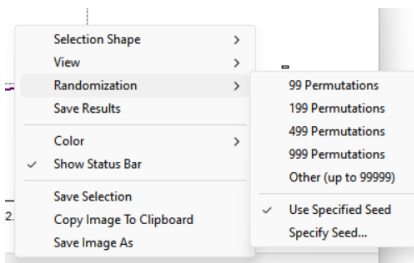
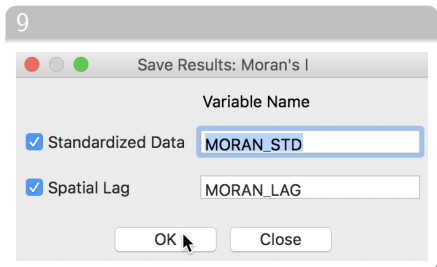(make sure that the tick mark for "Use Specified Seed" is kept)

The result of the permutation operation is a reference distribution for the statistic, depicted as a histogram. The green line shows the value of the statistic for the actual data.
What information can you draw from the distribution ?

Moran scatter plot options and Attribute Table

Right-click on the plot to bring up the available options. Save Results and View.

Verify the results by using the **Table** –> **Calculator** option to compute a standardized version of the Review Score and its spatial lag.

- Add a new variable called the standardized review score M_ST
- Make the variable constant as the review score
- Use the Univariate tab and the STANDARDIZED (Z) operator.

For the spatial lag, add M_LAG as the new variable and use the Spatial Lag tab, with the weights file created at the beginning and the previously standardized variable M_ST. Add a new variable called the standardized lag R_LAG to the data table.

Include the print screen of the final table in your analysis report.