Exercise 2

In Exercise 2 your task is to use desktop GIS software to analyze point pattern and polygon data using certain spatial analysis tools. The exercise is divided into three problems: Problem 1 (8 points), Problem 2 (11 points), and Problem 3 (1 point).

The instructions for solving the exercise are given for the ArcGIS Pro software. We will not provide instructions or support for using other tools for solving the exercise. Also be advised that the analysis functionality available varies between different software. Therefore, should you wish to use e.g. QGIS to solve the exercise, your results might be a bit different.

Due date

Due date for this exercise is on **Friday, October 24th.**

Time allocation

We expect that for most students solving this exercise will take between 5 and 10 hours of work.

What needs to be returned?

As an outcome of this exercise, you should write a short report that contains key visualizations of your work (screenshots are sufficient) and short descriptions of what they represent. In addition, the report should contain answers to all the questions that we ask related to the given problems/tasks. Please include intuitive headings to your report, such as “Problem 1 - Task 1” and indicate clearly which question you are answering, e.g. “Q1.1: My answer”.

Return the report into your personal Github repository as a single PDF file or a Word document. Name the file as “*exercise_2_MyGithubUserName.pdf*”

How to return the exercise to Github?

You can easily add a new file into your personal Github repository by:

1. Log into your Github account
2. Find and navigate to your personal exercise repository, such as *github.com/IntroSDA-2025/exercise-2-htenkanen*
3. Click “Add file” → “Upload files”
4. Drag and drop the report from your own computer into the repository
5. Add a short commit message (e.g. “Exercise 2 ready”)
6. Upload the file by clicking the “Commit changes” button

Hint: If you want to make changes to your report after you have submitted it to Github, you can simply upload the same (modified) file again using the **same filename**. Github will keep track of the changes to the file. Hence, you **don’t need to rename** the file when you have a different version of it.

Tools to be used

The analysis tasks in this exercise should be solved using a desktop GIS environment. We will provide you with instructions and support for solving the tasks using the ESRI ArcGIS Pro software.

Other tools are not required for solving this exercise.

Introduction

In this exercise you practice the use of the ArcGIS Pro desktop GIS system for analyzing and visualizing spatial data in vector format.

First, in Problem 1 you will perform a **Multi-Distance Spatial Cluster Analysis** (also known as *Ripley's K*) and a **Kernel Density analysis** to a point dataset. The data used in this exercise are building fire incident data for Helsinki from the years 2008 to 2010. The data is provided in text format as a CSV-file, and uses an old Finnish Coordinate System Kartastokoordinaatistojärjestelmä band 3 / Yhtenäiskoordinaatistojärjestelmä (KKJ3/YKJ). Thus, your task is to import the text format dataset to the desktop GIS system, assign the proper coordinate system and then apply the appropriate tools to create kernel density and Ripley's K analyses of the data.

In Problem 2 you will perform the **Spatial Autocorrellation** (global Moran's I) and **Cluster and Outlier Analysis** (Anselin Local Moran's I) analyses on **population density** in the Finnish population density dataset and the Paavo postal code area dataset. Similar to Exercise 1, the data sets for this analysis will be limited to those parts of the data that **intersect the Helsinki Capital Region** (cities of Helsinki, Espoo, Vantaa and Kauniainen). In this exercise it may be useful to have the input data as a data set and not just a layer extract from WFS data. Furthermore, for Paavo, you need to create a new field to hold the population density data.

Input data

Problem 1

Helsinki City building fires dataset from the years 2008 to 2010. Available in Github.

Problem 2

Population grid data for Finland (the same that was used in Exercise 1).
Paavo postal number area data (the same that was used in Exercise 1).

You may also need the Finnish municipality data to select the appropriate polygons from the input datasets.

ArcGIS functionality used

In order to solve the problems in this exercise, you need at least the following ArcGIS Pro functionality.

Problem 1:

XY Table to Point. Add the appropriate input table, the system will automatically find x and y coordinates. Select the appropriate Coordinate system (National Grids -> Finland -> Finland Zone 3, EPSG 2393).
https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/xy-table-to-point.htm

Multi-Distance Spatial Clustering Analysis. Number of distance bands can be adjusted. Do not decrease it from the default (10).
https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/multi-distance-spatial-cluster-analysis.htm

Kernel Density. Adjust Output Cell size and Search Radius appropriately. Both data use meters as units by default. In Output Cell Values -field use Densities as default, but you may also try out Expected Counts. Leave the Population field to NONE, since every point represents one building fire. Other input parameters do not need to be adjusted.
https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/kernel-density.htm

Problem 2:

Adding data and WFS connections. If you wish to get the input data from e.g. Paituli via WFS API instead of direct download.

Selection and creation of new layers from selected features. Only certain parts of the datasets are required in this exercise, since the area we're interested in is the Helsinki Capital Region. For Paavo the data needs to be exported into a new dataset since you need to calculate the population density. For the population grid it is sufficient to make a new layer in the map.

Attribute Field Calculator. The population density in the Paavo data needs to be calculated. This can be done by using he_vakiy (population) and pinta-ala (area in square meters) fields from Paavo data. Remember to create a new field and use float or double as the Field Type.
https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/calculate-field.htm

Spatial Autocorrelation (Global Moran's I). Select the proper input field (the one that represents population density) and select appropriate value for the Conceptualization of spatial relationships -parameter, and use row standardization. *Contiguity edges corners* represents 8-neighborhood (queen's move, Moore neighborhood). You can select "generate report" to get a bit more information about the results. After Moran's I has been calculated click on the View Details -link to see the results.
https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm

Cluster and outlier analysis (Anselin Local Moran's I). Select the proper Input Field (the one that represents population density) and select appropriate value for the Conceptualization of spatial relationships -parameter and use row standardization. *Contiguity edges corners* represents 8-neighborhood.
https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm

Problem 1 (8 points)

In Problem 1 you will do two different types of analysis on point pattern data. The first analysis is a **Multi-Distance Spatial Clustering analysis** or a *Ripley's K* analysis. Ripley's K is an analysis that shows whether a point dataset exhibits spatially significant clustering - or dispersion - at different distances. The second analysis is a **Kernel Density** analysis. Since the Kernel density is a *visual analysis method*, you need to experiment with it and use a number of different parameters to see what kind of heat maps are created with different parameterizations.

Task 1.1

In task 1.1 you make the Ripley's K and kernel density analyses.

To solve task 1.1 Do the following:

- Fetch data required for Problem 1, the Helsinki city building fires between 2008 and 2010. This data is available in Github.
- Use *XY table to point* too to create point data set
    - Remember to set the appropriate SRS for the data (Finland Zone 3, EPSG 2393)
- Calculate Multi-Distance Spatial Clustering (Ripley's K)
    - Do not set the weight field
    - See whether changing the Number of distance bands affects the results
- Calculate Kernel Density with various appropriate parameters
    - See how changing the following affects the results:
        - Output cell size (50-100 is a good starting value)
        - Search Radius (2000-3000 is a good starting value)
        - Output Cell Values (Densities or expected counts)
- Compare different outputs and interpret the results

In your answer provide the following:

1. **Images of the K-function charts for at least 2 different values for the Number of distance bands -parameters (1p)**
2. **Images of Kernel Density results with at least 2 different output cell sizes with the same search radius (1p)**
3. **Images of Kernel Density results with at least 3 different search radii with the same cell size (1p)**

Task 1.2

In task 1.2 you will analyze your results and answer the following questions. Write a few sentences about your thoughts as an answer for each question:

- **Question 1.1:** Interpret the results of Ripley's K analysis you did. What does the results say about the input data? (1p)
- **Question 1.2:** What does it mean if the ExpectedK is smaller than ObservedK? What about a case where ExpectedK is larger than ObservedK? (1p)

- **Question 1.3:** What happens to Kernel Density analysis when Search radius is changed? What about if the output cell size is changed? (1p)
- **Question 1.4:** Look at the raster values of the Kernel Density analysis results you made with different search radii. Are the analysis results comparable or not? (2p)

Problem 2 (11 points)

In Problem 2 you will analyze two different polygon datasets - the Finnish population grid and the Postal number areas - using both Global Moran's I and Anselin Local Moran's I. For both the attribute we're interested in is the *population density*. The ArcGIS Pro tools for this are the **Spatial Autocorrelation** and **Cluster and Outlier Analysis** tools. Both tools work on the spatial autocorrelation of the dataset, which in with the population grid and Paavo should be measured by comparing the population density of each polygon to its neighboring polygons.

Task 2.1

In task 2.1 you will analyze the data using Spatial Autocorrellation (Global Moran's I) and Cluster and Outlier Analysis (Anselin Local Moran's I).

To solve the task 2.1, do the following:

- Retrieve the population grid data and Paavo data
- Make a new layer from the population grid that intersects the Helsinki Capital Region (select appropriate polygons -> selection -> make layer from selected features)
- Make new data set from the parts of Paavo that intersect Helsinki Capital Region (select appropriate polygons, select data -> export features -> export selected features)
- Calculate the population density for Paavo data using Field Calculator. Remember to store the results in a new field, and use float or double as the field type
- Calculate Spatial Autocorrelation
    - Use the population density field as the input field
    - Select Contiguity edges corners as the Conceptualization of spatial relationships
    - Select Generate Report to get more detailed information about the results
    - After the analysis has been completed, click on View Details to see the Index value and open the more detailed report
- Calculate Cluster and Outlier Analysis
    - Adjust Input Field and Conceptualization of spatial relationships the same way as with Spatial Autocorrelation
    - Remember to check the scatterplot and histogram views in addition to the data in the map view
- Interpret the results

In your answer provide the following:

1. **The Moran's Index and z-score values for the two datasets (1p)**
2. **Images of map views of the Cluster and Outlier Analysis results for the two datasets (1p)**
3. **Images of Moran's Scatterplots for the two datasets (1p)**

Task 2.2

Assess the results of your analysis and answer to the following questions

- **Question 2.1:** What do the Moran's Index values for the datasets mean? Is the data spatially clustered or not? (1p)
- **Question 2.2:** What do the attribute values LMiZScore RS and LMiPValue RS in the Cluster and Outlier Analysis results layer mean? (1p)
- **Question 2.3:** What is the physical reason behind the high-low outlier polygon in the population grid dataset? (2p)
- **Question 2.4:** What is the overall interpretation of the Cluster analysis results? Is this conclusion realistic? (2p)
- **Question 2.5:** Are there significant differences between the Moran's I results for the population density grid and the postal area dataset? (2p)

Problem 3 (1 point)

To help us to develop the exercises, and understand the workload for you to complete the Exercise, **please provide an estimate of how many hours you spent doing this exercise?**

In addition, if you would like to give any feedback about the exercise, you can add comments under the Problem 3 (optional).