

Exercise 9

Spatial relative risk analysis (SPARR)

Overview

In this exercise, you will explore spatial patterns of relative risk for different BMI categories within a study area using R. The analysis includes both adjusted and unadjusted BMI categories (Obese, Overweight, Normal weight, Underweight), computation of asymptotic p-values, Monte Carlo permutations, and visualization in QGIS.

Setup Instructions

Before starting the exercise, make sure you have the required software and R packages installed. Follow the instructions [here](#) to download R and RStudio for your operating system.

- Open **RStudio** and type the following command in the Console to install the required packages: `install.packages(c("sparr", "spatstat", "sf", "dplyr", "raster"))`
- **Verify Installation:** Once the installation is complete, run the following command to ensure the packages are working correctly: `library(sparr) library(spatstat) library(sf) library(dplyr) library(raster)`

SPARR is a tool designed for analyzing binary variables, allowing you to identify regions of higher or lower relative risk. In your project, you can use SPARR to explore any binary variable of interest and detect spatial patterns related to your research question.

For more details on the SPARR package, you can refer to the [SPARR Documentation](#).

Data and Context

You are provided with geospatial health data for individuals within Lausanne. The data includes:

- **BMI Categories:**
 - Obese: Unadjusted obesity indicator (1 = obese, 0 = not obese).
 - Adj_Obese: Adjusted obesity indicator (adjusted for median income).
 - Similar indicators are provided for Overweight, Normal weight, and Underweight.
- **Spatial Data:**
 - data_sparr.gpkg: Geopackage containing individual health data with coordinates.
 - Lausanne_Extent.shp: Shapefile defining the study area boundary.

Instructions

Part 1: Data preparation and function understanding

1. **Load the R Code :**

Exploratory data analysis in environmental health

Dr Stéphane Joost, Dr Mayssam Nehme, Noé Fellay

- Open the provided SparrCode.R. Follow the code step-by-step by running one cell at a time using **Shift + Enter**.
 - The `analyze_bmi_category` function calculates the relative risk surface for a given BMI category. It allows for fixed or adaptive bandwidths (`adapt = TRUE` or `FALSE`).
2. **Asymptotic p-values explanation:**
 1. If `tolerate = TRUE`, the function computes asymptotic p-values, indicating the level of significance for increased risk using tolerance contours ([Hazelton and Davies, 2009](#); [Davies and Hazelton, 2010](#)).
 2. Smaller p-values represent stronger evidence of a significantly increased risk.
 3. **Load the data:**
 - Use the cells provided to load the data for analysis.

Part 2: Analyzing BMI categories

1. **Run the analysis for "Obese":**
 - Use the `analyze_bmi_category` function with different fixed bandwidths (`adapt = FALSE`). What differences do you observe based on the chosen bandwidth? What is the optimal bandwidth chosen by the oversmoothing method?
2. **Adjusted obesity analysis:**
 - Analyze the "Adj_Obese" category. Are there differences between obesity and obesity adjusted for income? What might explain these differences or lack thereof?
 - Run the function with an adaptive bandwidth (`adapt = TRUE`) without specifying a bandwidth. What is the difference between fixed and adaptive bandwidths ?
3. **Other BMI categories:**
 - Perform the analysis for overweight, normal weight and underweight with a 300-meter fixed bandwidth. What spatial patterns of relative risk do you observe for each category? Are there any differences between adjusted and unadjusted values?

Part 3: Significance testing with Monte Carlo permutations

1. **Data preparation:**
 - Use the provided code to prepare data for **the variable of your choice**. (line 143)
2. **Monte Carlo permutations:**
 - Compute the relative risk surface with your choice of **bandwidth** and decide if you'd like to use a **fixed or adaptive bandwidth** by setting `adapt = TRUE` or `FALSE`.

Exploratory data analysis in environmental health

Dr Stéphane Joost, Dr Mayssam Nehme, Noé Fellay

- **Note:** If you set `adapt = TRUE`, the computation will take significantly longer, especially with a high number of permutations. Consider the following tips:
 - **For slower computers:** Use `adapt = FALSE` for faster processing.
 - **If computation time is too long:** Reduce the number of permutations (e.g., use 500 instead of 999) and indicate the number of permutations used on your final map.
 - **Recommendation:** For exploratory analysis, start with a lower number of permutations if `adapt = TRUE`.

- Describe how Monte Carlo permutations work.

3. Plot the results:

- Display the relative risk surface with contours at significance levels of 0.01 and 0.05 with `test="upper"`. Consider using `test = "two-sided"` to explore both high and low risk or `test = "lower"` for lower risk only.

Part 4: Mapping and visualization with a basemap

1. Export Data:

- Save the relative risk surface and p-value surface for adjusted obesity as GeoTIFF files.

2. Import into QGIS:

- Load `relative_risk_surface.tif` and `pvalue_surface.tif`.
- Add a basemap like **OpenStreetMap**.

3. Create the map:

- Visualize the relative risk surface using a color ramp like "RdBu":
 - Red for high relative risk of obesity.
 - Blue for low relative risk.

4. Overlay P-value contours :

- Convert the p-value raster to polygons and select the 0.01 and 0.05 significance levels.
- Dissolve the polygons so that you have 2 polygons, 1 that represents p-values below 0.01 and other p-values below 0.05.
- Smooth both polygons using:
 - **Processing Toolbox** > Smooth, with **10 iterations** and **offset = 0.4**.

5. Finalize the Map:

Exploratory data analysis in environmental health

Dr Stéphane Joost, Dr Mayssam Nehme, Noé Fellay

- The map should contain the legend (i.e. the relative risk p-values at 0.01 and 0.05 levels, as well as the number of obese and other individuals, a title, a scalebar, a north arrow and sources.

Discuss with your project group: Could a similar analysis in Geneva, using a binary variable of your choice and a custom bandwidth, help answer a research question for your final project? If so, how would you apply it?