

MM928 Project

Project Overview

In this project you will explore a variety of spatial data with the goal of assessing the impact of air pollution on admissions to hospital due to respiratory illness across Glasgow. You will be assigned a year of data and a specific pollutant to examine which should be subsetting from the data sets provided.

The pollutant data are modelled quantities provided by DEFRA. These are stored in a csv file called `Pollutants.csv`. These represent values at 340 locations in and around the city of Glasgow.

The admissions data is stored in a csv file called `Admissions.csv`. This includes data on observed counts of respiratory admissions in a column `Respiratory.Prim` along with expected counts and measures of deprivation which should be considered as confounding variables.

The shapefile is stored in the folder `Shapefiles`.

Part 1: Estimating Pollutants (20 marks)

The goal of this analysis is to use the pollutant values in a model for respiratory admissions. To do this we need to have an estimate of pollution in each intermediate zone in Glasgow. In this section you should:

- Explore the spatial distribution of pollutants
- Examine the sample variogram and fit an appropriate variogram model
- Compare different methods for spatial prediction
- Use the selected method to predict on each intermediate zone

To get the coordinates for the final prediction you will need to have read in your shapefile, and use the `coordinates()` function on the spatial object.

Part 2: Modelling (30 marks)

In this part of the project you will build models for respiratory admissions using the pollutant information obtained in part 1 and the other covariates provided. In this part of the project you should:

- Combine pollutant data, admissions data and shapefile into one object
- Explore the spatial pattern in respiratory admissions

- Explore relationships between each potential covariate and respiratory admissions
- Fit Poisson Regression and Poisson CAR models
- Examine convergence of models
- Assess model assumptions
- Interpret the estimated pollutant effect based on the best model

Presentation of report, clarity of language (50 marks)

Total (100 marks)

Report Structure, content and submission

The report itself does not need to follow any particular structure. For example, separate sections that briefly document the work of *Part 1* and *Part 2* are fine. However, you should ensure that any analysis you carry out is clearly interpreted using full sentences. You should assume that the reader is familiar with the methods, but should clearly explain any decisions that you are making based on your analysis.

- The report should be a maximum of 10 pages in length including graphics and tables
- Graphs should be suitably labelled, sensibly scaled and cropped
- Numerical R outputs should be neatly presented in tables or presented in the text
- Your R code should be included as an appendix to your report - there should be no R code included in the text