Spatial Statistics Project

1 Project overview

In this project you will estimate the effects of air pollution and socio-economic variables on the risk of admission to hospital caused by respiratory disease. The data for the project cover the 7 year period spanning 2003 to 2009 and describe annual hospital admissions across 624 electoral wards of Greater London. Each person will be individually assigned both a year and a data set to work with. The data are contained in the file `london.RData`, and once loaded, the data are in the form of an `sp` object called `london` with geographical attributes included. There are different versions of the data and each person has been allocated a different version of the data. Further information about the contents of the data can be found in Section 4.

2 Project tasks

1. (a) Create a subset of your assigned data for the year you have been assigned.
   (b) Calculate the standardised incidence ratio (SIR) for respiratory disease and add this as a new column to your data subset.
   (c) Use a spatial plot to explore SIR.
   (d) Use scatterplots to explore the association between SIR and `PM25`, `JSA` and `Price`.

   (10 marks)

2. (a) Use `BUGS` or `Nimble` and appropriate prior distributions to fit a Poisson regression model for respiratory disease, making use of the columns `Observed`, `Expected`, `PM25`, `JSA` and `Price`.
   (b) Use the Gelman-Rubin diagnostic (`Rhat`) and Geweke diagnostic plots to assess whether convergence has been reached.
   (c) Calculate the Pearson residuals and use these to check the model assumptions.

   (10 marks)

3. (a) Extend the model developed in 2. by including spatial random effects ϕ with a Conditional-autoregressive prior distribution.
   (b) Use the Gelman-Rubin diagnostic (`Rhat`) and Geweke diagnostic plots to assess whether convergence has been reached.
   (c) Calculate the Pearson residuals and use these to check the model assumptions.

   (10 marks)

4. Construct a table with columns for DIC, Moran's I statistic, variance of the Pearson residuals and the estimated effect of `PM25` (and 95% credible interval) to critically compare the models in 2. and 3. Using estimates from the better model, interpret the effect of `PM25` on the rate of respiratory hospital admission in Greater London.

   (10 marks)

3 Report structure, content and submission

The report itself does not need to follow any particular structure, however, you should ensure that any analysis you carry out is **clearly interpreted**, using full sentences. You should write as though your audience were your statistics classmates who are unfamiliar with the data.

- The report should have a cover page with your name and student ID clearly marked
- The report should be between 4 and 6 A4 pages in length including graphics and tables but excluding cover page and any references
- Graphs should be suitably labelled, sensibly scaled and cropped
- Numerical `R` outputs used to answer questions should be neatly presented in tables or in the text
- Your `R` code and `BUGS` code (if appropriate) should be included in an appendix
- **Please submit your report by midday on Friday 8th April via the upload link on Myplace.**

4 Data description - `london.RData`

Column name	Description
`Observed`	Annual hospital admissions due to respiratory disease
`Expected`	Expected annual hospital admissions due to respiratory disease based on national average rate
`JSA`	Proportion of electoral ward in receipt of jobseeker's allowance, an unemployment benefit
`Price`	Annual average sale price of homes (logged GBP)
`PM25`	Annual average concentration of fine particulate matter $PM_{2.5}$ ($\mu g l^{-1}$)
`year`	Year of observation: 2003 to 2009 inclusive