Exercise sheet 3 – Spatial statistics for epidemiologists October 27-29, 2010
Leonhard Held, Birgit Schroedle Date: Thursday, October 28, 2010

Exercise 1 – Defining a binary indicator for the distance to a health hazard

In 1992, a survey was conducted to investigate the association between asthma and four possible health hazards: Source 1 is a coking works, source 2 a chemical plant, source 3 a waste treatment center and source 4 the main road network of the study region. 2133 school children were included in the study and labelled as cases and controls, respectively. It should be assessed, if the distance between the residence location of a child and the chemical plant has an influence on the risk to get asthma.

- Load the libraries `splancs`, `fields` and `vcd`. Load the dataset `asthma.RData.txt`. Look at the variables in the dataset using `names()`, extract cases and controls from the dataset and plot them.
- Chose an appropriate threshold (e.g. 0.5) for the distance to the chemical plant (`d2source2B`), define a binary indicator and plot a contingency table. Calculate the odds ratio by hand. Which value of the odds ratio would mean that the distance to the chemical plant doesn't have an influence on childhood asthma?
- Print the odds ratio and the respective p-value using the call from the lecture notes. Do you get the same results as for the distance to the coking works?

Exercise 2 – Fitting a binomial GLM including distance to a health hazard to spatial case-control data

The `asthma` dataset contains several covariates which could be confounding the influence of distance of residential locations from the chemical plant. To adjust for them a binomial GLM has to be fitted.

- `attach()` the dataset `asthma`. Fit a binomial GLM including just the binary distance indicator and compare your results to the results of Exercise 1.
- Identify a significant covariate that is different from those found in the lecture. You can also use the AIC to compare the models. Interpret the obtained coefficients.

Exercise 3 – Using a nonlinear function to model the distance from a health hazard

It was suggested by Diggle and Rowlingson (1994) that a nonlinear function should be used to model the influence of the distance to a health hazard. This methodology is implemented in the function `tribbleNG()` in the `tribble2` package (Musti et al., 2009). Since a numerical approximation algorithm is used, initial values for all unknown parameters have to be provided.

- Load the `R` package `tribble2`. Read through `?tribbleNG()`.
- Check the significance of the distance from the coking works (`d2source1B`) and the chemical plant (`d2source2B`). The needed function call for the coking works, including initial values for all unknown parameters, is
```
source1<-tribbleNG(ccflag=Asthma,rho=sum(Asthma==1)/sum(Asthma==0),
                                  distances=d2source1B,alpha=1,beta=1).
```
Verify the result for the coking works using the lecture notes.
- The next step is to include covariates into the model. Check the significance of the covariates `CAllergies` and `HayFever`. It is best to use the estimates from the model `source1` as initial values for α, β and ρ now.

Diggle, P and Rowlingson, B (1994): A conditional approach to point process modelling of elevated risk, *Journal of the Royal Statistical Society, Series A*, **157(3)**, pp. 433-440.

Musti, M et al. (2009): The relationship between malignant mesothelioma and an asbestos cement plant environmental risk: a spatial case-control study in the city of Bari (Italy), *International archives of occupational and environmental health*, **82(4)**, pp. 489-497.

Exercise 4 – Calculate the ratio of two kernel estimates to get an adjusted risk surface

In the years 1974-1983 locations of cases of lung and larynx cancer in Chorley-Ribble, Lancashire, were recorded.

- Load the library `splancs` and the dataset `southlancs`.
- Extract the larynx and lung cancer cases from the dataset and plot their locations.
- Use a `2d` kernel smoother to estimate the continuous risk surface of both diseases separately (see Exercise sheet 1, Exercise 1), plot them using `filled.contour()` and save these plots using `pdf()`.
- Estimate the ratio of the two kernels using the function `kernrat()` and save this plot also in pdf format. Compare the three plots.