2020 STAT 555
Statistical Methods for Spatial Epidemiology
Assignment 2

To be submitted to the canvas site by the start of class on Wednesday 5th February, 2020.

Hand in your `R` code as an appendix.

In this question we will carry out disease mapping for Ohio lung cancer mortality data from 1988. On the class website is a data file that contains, for each county, the observed deaths and expected deaths, along with the polygon files.

Let Y_i and E_i, $i = 1, \dots, n$, denote the observed and expected counts in county i, $i = 1, \dots, n$. Then consider the model

$$Y_i | \theta_i \sim \text{Poisson}(E_i \theta_i). \tag{1}$$

The expected numbers are adjusted for gender (0=male, 1=female), race (0=white, 1=non-white) and age (7=45–54, 8=55–64, 9=65–74, 10=75–84), which are obtained using internal standardization[1]

1. (a) Provide a map of the observed counts Y_i.
   (b) Provide a map of the expected counts E_i.
   (c) Provide a map of the SMRs, defined as

   $$\text{SMR}_i = \widehat{\theta}_i = \frac{Y_i}{E_i},$$

   for $i = 1, \dots, n$. Comment on the variability of the SMRs.

2. In this question we will smooth the SMRs using the simple disease mapping Poisson-Gamma model

$$\begin{aligned} Y_i | \beta_0, \delta_i &\sim_{ind} \text{Poisson}\left(E_i \text{e}^{\beta_0} \delta_i\right) \\ \delta_i | \alpha &\sim_{iid} \text{Ga}(\alpha, \alpha), \end{aligned}$$

[1] Let Y_{ij} and N_{ij} denote the disease counts and populations in county i and strata j, $i = 1, \dots, n$ and $j = 1, \dots, J$, where n is the number of counties and J the number of strata. Then reference risks obtained by internal standardization correspond to $p_j = \sum_{i=1}^n Y_{ij} / \sum_{i=1}^n N_{ij}$. The expected numbers are then $E_i = \sum_{j=1}^J N_{ij} p_j$.

for i, $i = 1, \ldots, n$.

(a) Use the `eBayes` function in the `SpatialEpi` package to obtain empirical Bayes estimates of $\text{RR}_i = \text{e}^{\beta_0}\delta_i$ (these are the posterior means, and are the quantities named `RR` in the object created from the call to `eBayes`). Map these estimates. Plot the empirical Bayes estimates against the SMRs and comment.

(b) Use the `EBpostdens` function in the `SpatialEpi` package to plot the gamma posterior densities of the relative risks, for $i = 1, \ldots, 4$.

(c) Using the function `EBpostthresh` in the `SpatialEpi` package, calculate the posterior probabilities that RR_i exceeds the threshold 1.2, and map these probabilities. Are there regions in which the relative risk appears significantly high?

3. In this question we will smooth the SMRs using the disease mapping Poisson-Lognormal model:

$$\begin{aligned} Y_i|\beta_0, \epsilon_i &\sim_{ind} \text{Poisson}(E_i \text{e}^{\beta_0}\text{e}^{e_i}) \\ e_i|\sigma_e^2 &\sim_{iid} \text{N}(0, \sigma_e^2) \end{aligned}$$

for i, $i = 1, \ldots, n$.

(a) Using the `inla` function in `R` fit this model using the default priors for β_0 and σ_e. Report the posterior medians and 95% intervals for β_0 and for σ_e.

(b) Extract the relative risk estimates and provide a map of these. Compare these estimates with the SMRs and with those obtained from the Poisson-Gamma model.

4. **Bonus Question:** Suppose that instead of having available the counts and expected numbers we have access to the relative risks $\widehat{\theta}_i = Y_i/E_i$ and their standard errors $\sigma_{\epsilon i} = \sqrt{\widehat{\theta}_i/E_i}$.

Take the data as $Z_i = \log\widehat{\theta}_i$ and fit the model

$$Z_i = \log\widehat{\theta}_i \sim \text{N}(\beta_0 + e_i, \sigma_{\epsilon i}^{\star 2}),$$

where $\sigma_{\epsilon i}^{\star 2} = \sigma_{\epsilon i}^2/\widehat{\theta}_i^2 = 1/(E_i\widehat{\theta}_i)$. We then place smoothing priors on via $e_i \sim_{iid} \text{N}(0, \sigma_e^2)$.

Fit this model using `inla` and plot the estimated relative risks from this model against the estimates from the Poisson-Lognormal model, and comment.