2025 STAT 554/CSSS 554/SOC 534
Statistical Methods for Spatial Data
Assignment 3

To be submitted to the canvas site by midnight on Wednesday 5th February, 2025.

Hand in your `R` code as an Appendix.

In this question we will carry out disease mapping for lung cancer mortality data for men in the Valencia region of Spain from 1991-2000. For more details on the data, see Martinez-Beneito et al. (2019, Disease Mapping). On the class website is a data file that contains, for each subarea: observed deaths and expected deaths (adjusted for reference rate only), polygon files, and a graph file for INLA.

Use the following code to read in the data:

```
# Set your working directory to where HW3data.Rdata and VR.graph are located
load(HW3data.Rdata)
# Expected counts are stored in Exp.mv3[,"Lung"]
# Observed counts are stored in Obs.mv3[,"Lung"]
# The shapefile for plotting is VR.cart
```

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

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

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

(d) Plot the SMRs versus the estimated standard errors, which are given by $\sqrt{\widehat{\theta}_i/E_i}$.

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

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

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 posterior medians of the relative risk (RR) estimates and provide a map of these.

(c) Plot these posterior RR estimates against the SMRs, and comment.

(d) Plot the posterior standard deviations of the RRs against the standard errors of the SMRs and comment.

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

$$\begin{aligned} Y_i|\beta_0, S_i, \epsilon_i &\sim_{iid} \text{Poisson}(E_i \mathrm{e}^{\beta_0} \mathrm{e}^{S_i+\epsilon_i}) \\ \epsilon_i|\sigma_\epsilon^2 &\sim_{iid} \mathrm{N}(0, \sigma_\epsilon^2) \\ S_1, \ldots, S_n|\sigma_s^2 &\sim \text{ICAR}(\sigma_s^2) \end{aligned}$$

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

(a) Using the `inla` function in `R` fit this model using the `bym2` model, with the default prior for β_0 and the following prior specification for the spatial and non-spatial random effects (note that you must be in the directory that contains the `VR.graph` file):

```
f(Region, model="bym2", graph="VR.graph", scale.model=T, constr=T,
hyper=list(phi=list(prior="pc", param=c(0.5, 0.5), initial=1),
prec=list(prior="pc.prec", param=c(0.3,0.01), initial=5)))
```

These choices correspond to the prior belief that there is a 1% chance that the total residual standard deviation is greater than 0.3, and a 50% chance that the proportion of the variance that is spatial is bigger than 0.5.

Report both the posterior medians and 95% intervals for β_0, the total variance of the random effects, and the proportion of the total variance attributed to the spatial random effect.

(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-Lognormal model (i.e., the model with IID random effects only) that you fit in Question 2.

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

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

   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 $e_i \sim_{iid} \mathsf{N}(0, \sigma_e^2)$.

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