2022 STAT 554/CSSS 554/SOC 534
Statistical Methods for Spatial Data
Assignment 3

To be submitted to the canvas site by the start of class on Wednesday 22nd February, 2023.

Hand in your `R` code as an appendix.

In this question we will carry out SAE for smoking prevalence in health reporting areas (HRAs) in King County, using the BRFSS data. The smoking variable is labeled `smoker1`. The "554-2023-SUMMER.pdf" notes will be helpful for carrying out this assignment. Throughout this assignment, for the Bayesian analyses, use the default `INLA` hyperpriors.

In the following, we let Y_i and n_i denote the number of smokers and the number sampled respectively, and define p_i to be the proportion of smokers in HRA i, $i = 1, \ldots, n$.

1. **Naive Estimation:** As a naive first analysis, consider the binomial model:

   $$Y_i|p_i \sim_{iid} \text{Binomial}(n_i, p_i).$$

   Calculate the sample proportions $\widehat{p}_i = y_i/n_i$, and the standard errors $\sqrt{\widehat{p}_i(1-\widehat{p}_i)/n_i}$. Map the estimates and standard errors.

2. **Weighted Estimation:** Create a `svydesign` object and calculate weighted estimates $\widehat{p}_i^w$, with associated standard errors and map each of these.

3. Plot the naive and weighted estimates of p_i against each other and comment. Plot the naive and weighted standard errors of p_i against each other and comment. Are the naive or the weighted the most appropriate summaries? Why?

4. **Smoothed Naive Estimation:** Now we consider a simple binomial smoothing model:

   $$\begin{aligned} Y_i|p_i &\sim_{iid} \text{Binomial}(n_i, p_i) \\ \theta_i = \log\left(\frac{p_i}{1-p_i}\right) &= \alpha + b_i, \end{aligned}$$

   where α is the intercept and b_i are BYM2 random effects. Fit this model using `INLA` and extract posterior medians and posterior standard deviations of p_i. Map these quantities.

5. Plot the naive and smoothed binomial estimates of p_i against each other and comment. Plot the naive and smoothed binomial standard errors of p_i against each other and comment.

6. **Smoothed Weighted Estimation:** Define $\widehat{\theta}_i = \log[\widehat{p}_i^w/(1-\widehat{p}_i^w)]$ to be the transformed weighted estimates and $\widehat{V}_i$ to be the estimated design-based variances of $\widehat{\theta}_i$. Fit the model

$$\begin{aligned} \widehat{\theta}_i|\theta_i &\sim_{iid} \mathsf{N}(\theta_i, \widehat{V}_i) \\ \theta_i &= \alpha + b_i, \end{aligned}$$

   where α is the intercept and b_i are BYM2 random effects. Fit this model using the `SUMMER` package and extract posterior medians and posterior standard deviations of p_i. Map these quantities.

7. Plot the weighted and smoothed weighted estimates of p_i against each other and comment. Plot the weighted and smoothed weighted standard errors of p_i against each other and comment. Which of the weighted or the smoothed weighted would you recommend using? Why?

8. Summarize the HRA variation in smoking prevalence across King County.