

STATS 506 Project Report

Zicong Xiao, Yulin Gao, Qianang Chen, Xiaoyang Sheng

December 7, 2022

1 Introduction

1.1 Topic

An urban heat island (UHI) is an urban area which becomes islands of higher temperatures relative to the surrounding rural areas. This phenomenon is caused by human activity, materials of infrastructures, climate change and etc. It will cause residents some severe diseases like asthma and stroke, especially to population who are over 65 years old. Yet the government can adopt some measures, for example, increase the insurance rates and medical expenditure to mitigate the harmful effects. We looked into data from mainly two sources: First we collected the climate data from earth engine, with covariates like surface temperature and precipitation, which are significant factors in our model. On the other hand we extracted the health data from CDC and chose the variables like mortality of different age groups, insurance rates and prevalence of asthma. Then we'll use different models (linear regression, Bayesian hierarchical model, random forest) to make vulnerability assessment towards the health of older adults. Next we'll do some model checking and validation to evaluate the model fitness. Finally based on the results we obtained, we'll do some analysis on negative impact of UHI on public health, along with the effect of each covariate.

2 Investigation Plan

2.1 Data set

- United States Surface Urban Heat Island database; from Mendeley Data: <https://data.mendeley.com/datasets/x9mv4krnm2/2>
- Global map temperature; from NASA: <https://giovanni.gsfc.nasa.gov/giovanni/#service=TmAvMp&starttime=&endtime=&variableFacets=dataFieldMeasurement%3ASurface%20Temperature%3B&dataKeyword=AIRS3STM>
- Death of elderly people by county from CDC: <https://wonder.cdc.gov/ucd-icd10.html>
- Insurance rate of each US county from US Census: <http://data.ctdata.org/dataset/health-insurance-coverage>

3 Bayesian Linear Model

3.1 Model Selection

In this section we build the Bayesian linear model with respect to the prevalence of stroke and UHI/climate effects, and the insurance rate is also taken into account. Let $y_i (i = 1, 2, \dots, N)$ be the mean of people who had got stroke within a county in three years (2017-2019), $N = 427$ is the number of counties we choose. We define x_1 as the UHI index in summer during the day, x_2 as the UHI index in summer at night, x_3 as the mean air temperature and x_4 as the insurance rate within a county. The likelihood distribution is defined as follows:

$$y_i \stackrel{\text{iid}}{\sim} N(\mu_i, \sigma^2) \quad (1)$$

$$\mu_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (2)$$

Then we define the noninformative priors of parameters as follows:

$$\begin{aligned} \beta_i &\sim U(-1e6, 1e6), i = 0, 1, 2, 3, 4 \\ \sigma^{-2} &\sim G(-1e6, 1e6) \end{aligned} \quad (3)$$

3.2 Fitting and Results

We use flat priors, 5 chains, 100K iterations and 50K burn-in periods in our MCMC simulation. The posteriors of each parameter of population over 65 and under 65 are as follows:

Parameter	mean(over 65)	95% CI(over 65)	mean(under 65)	95% CI(under 65)
β_0	-261.758	(-803.703,173.850)	-168.118	(-219.075,-99.355)
β_1	5.135	(-0.028,10.241)	1.258	(0.595,1.920)
β_2	1.956	(-10.971,15.151)	1.010	(-0.630,2.643)
β_3	1.960	(0.772,3.502)	0.703	(0.517,0.840)
β_4	-0.557	(-1.899,0.884)	-0.174	(-0.360,0.001)

3.3 Analysis and Discussion

From the tables above we find that the effect of UHI index is more significant than mean air temperature and the prevalence of stroke in all age groups will grow with the increase of UHI index and mean air temperature.

We also find that although β_2 is positive, but the 95% credible intervals both contain 0. Use R, the probabilities that $\beta_2 > 0$ from each group are 0.6136 and 0.8854, so we can conclude that the UHI index in summer at night has weaker effect to public health than the other two covariates. The UHI effect is stronger at night based on scientific literatures, which means the index tend to be higher. But the temperature at night is also lower, so people aren't as vulnerable to heat-related illness, which causes $\beta_2 < \beta_1$.

The mean value of β_4 is negative in both tables. If the rate of insurance coverage within a county is higher, which depends on government funding and people's awareness of health, then the residents will be less vulnerable to all kinds of diseases.

The first three parameters in table 1 are all greater than the corresponding parameters in table 2, so the variables of mean air temperatures and UHI index have a greater effect to older adults who are 65 years old. Since their immunities are going down, they are more susceptible to external factors than young adults. The parameter β_4 in table 1 is less than that in table 2, which implies that the insurance coverage guarantee more interests to older populations, and their diseases may be more treatable.

We can improve the fitness of model by increasing the number of chains, burn-in periods and thinning rates, or considering the interaction between variables.

4 Appendix

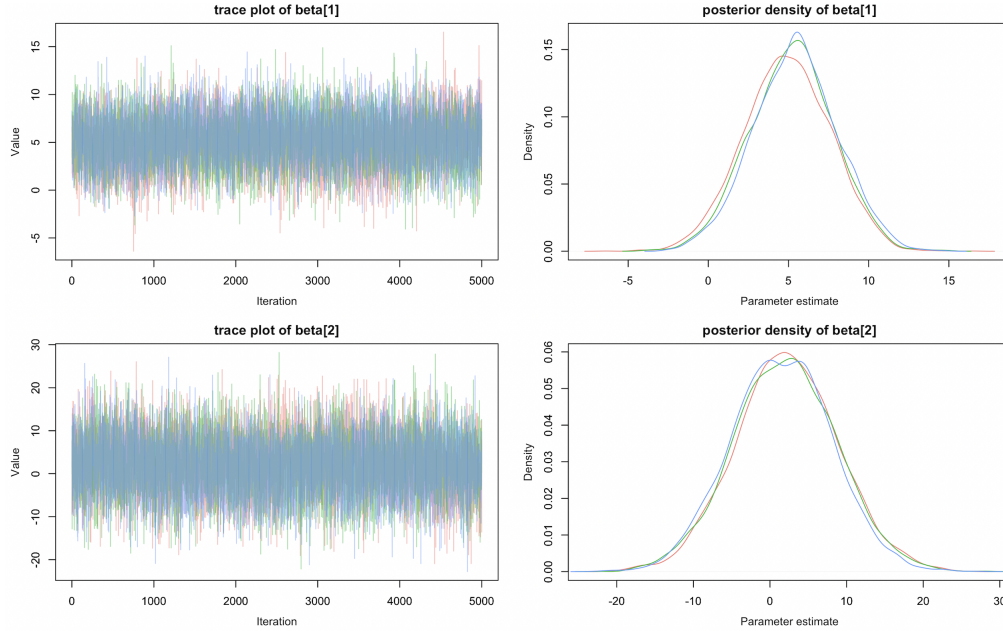


Figure 1: The trace plot and posterior pdf of β_1 and β_2 from the MCMC output of population over 65

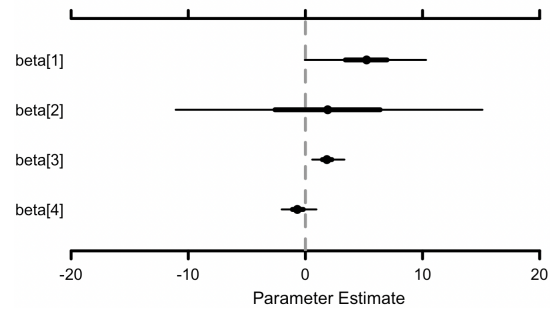


Figure 2: The caterpillar plot of β_i from the output of population over 65

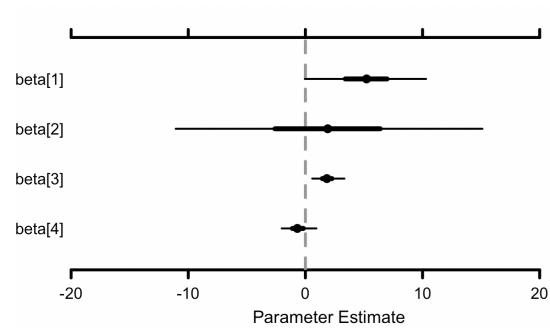


Figure 3: The caterpillar plot of β_i from the output of population under 65