BST 6200

Homework 4

Due 5:00 pm CDT May 4, 2020

Overall Goal: Perform an ecological study of the relationship between smoking and pancreatic cancer rates in the state of Minnesota.

Obtain the shape file for Minnesota through the tigris package.

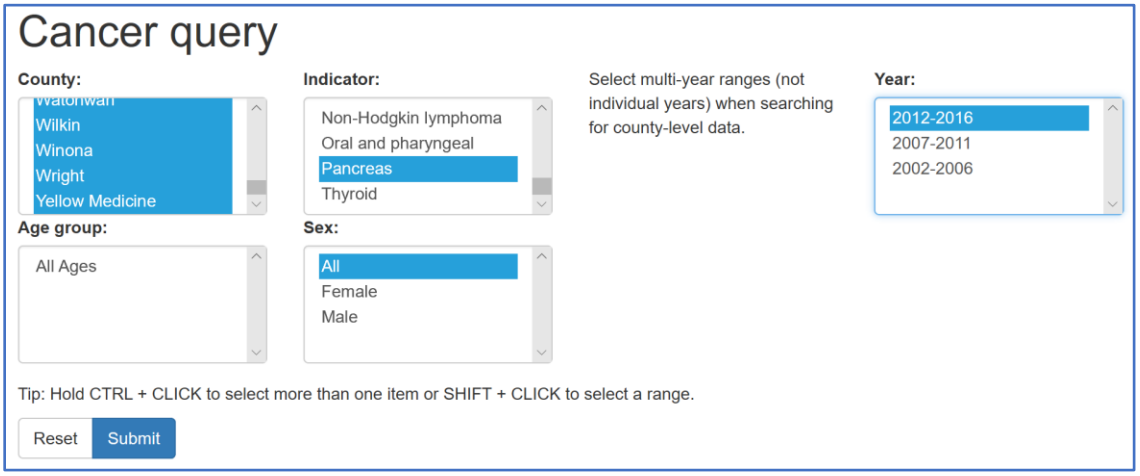
Obtain the smoking data by selecting Minnesota and Adult Smoking here:

https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/tobacco-use

You can probably copy the data and paste it into Excel.

Obtain the number of cases and the population size here:

https://data.web.health.state.mn.us/cancer_query



Use shift-click to select all counties. Then select "Pancreas" under Indicator, "2012-2016" under Year, and "All" under Sex. Then click "Submit" and finally "Download" at the bottom of the data. Two of the columns in the resulting file are "count" and "population". You will need both of these. You might want to clean up this file using Excel.

As part of your report, give choropleth maps of smoking rates and pancreatic cancer rates. Compute Moran's I for both and assess their significance.

Run a regression model of the form:

$$Y_i \sim POI(P_i \eta_i)$$

where

$$\eta_i = \exp(\beta_0 + \beta_1 x_i + u_i + v_i)$$

$P_i =$ population of county i

u_i is correlated heterogeneity of county i

v_i is uncorrelated heterogeneity of county i

x_i is the smoking rate of county i

Write a report addressing the question of smoking and pancreatic cancer. The project shouldn't be too long. I'm thinking something like 3 to 6 pages, counting figures.

[Note: This project is much like the term project for the course, except I have pointed you to the data and given you a specific model with just one predictor variable (smoking). You should look for more than one predictor variable in your term project model.]