Patrick Liu

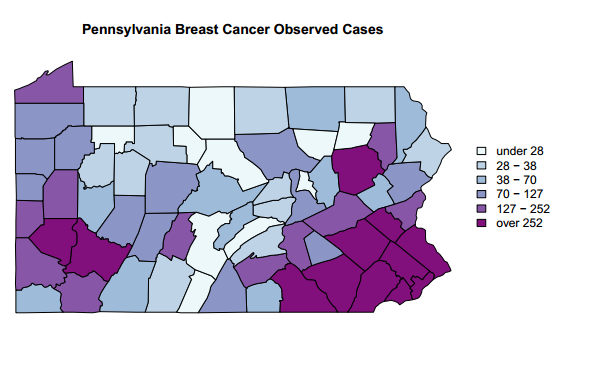
CS&SS554

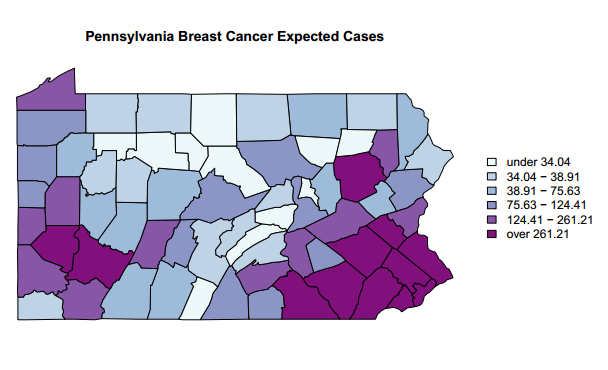
Homework 2

1/31/2017

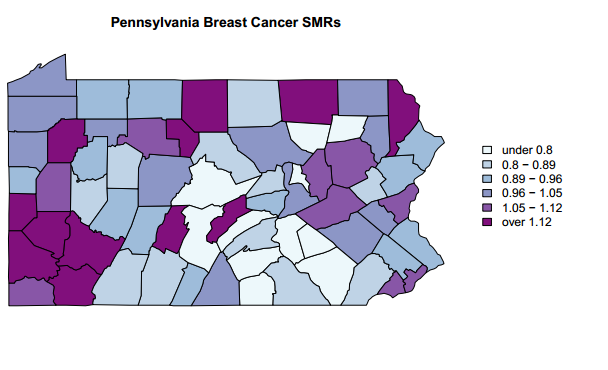
1.) is the relative risk (or SMR) for area i of breast cancer incidence. It is the ratio between the observed number of cases in area i and the number of cases that would be expected based on the standardized population in Pennsylvania.

2 .)

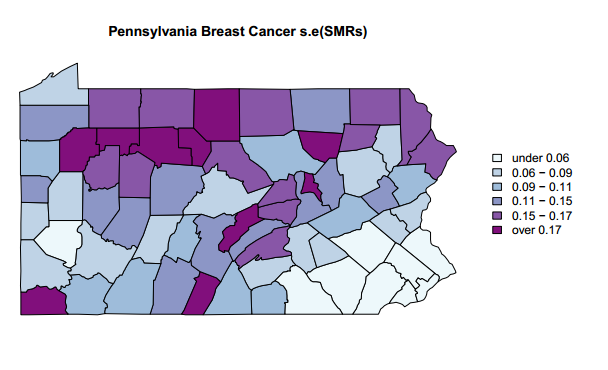




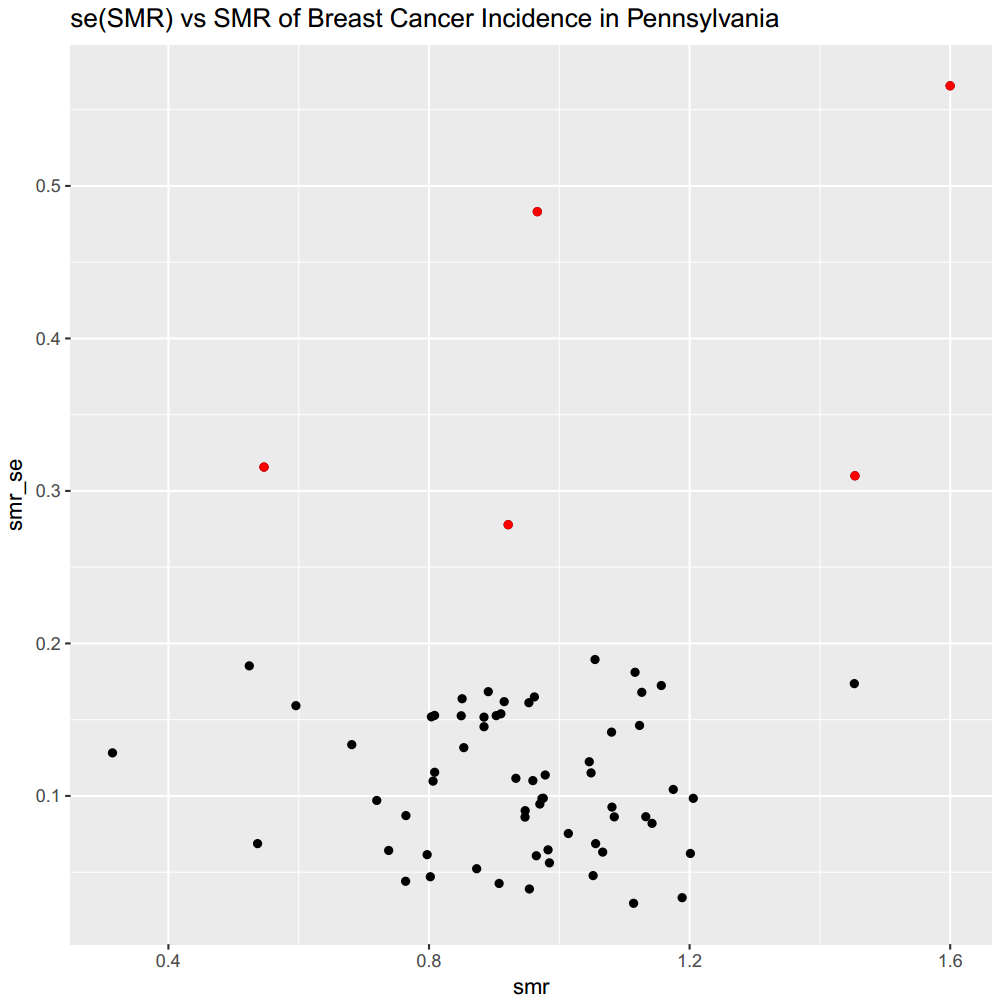
3.)



4.)



5.)

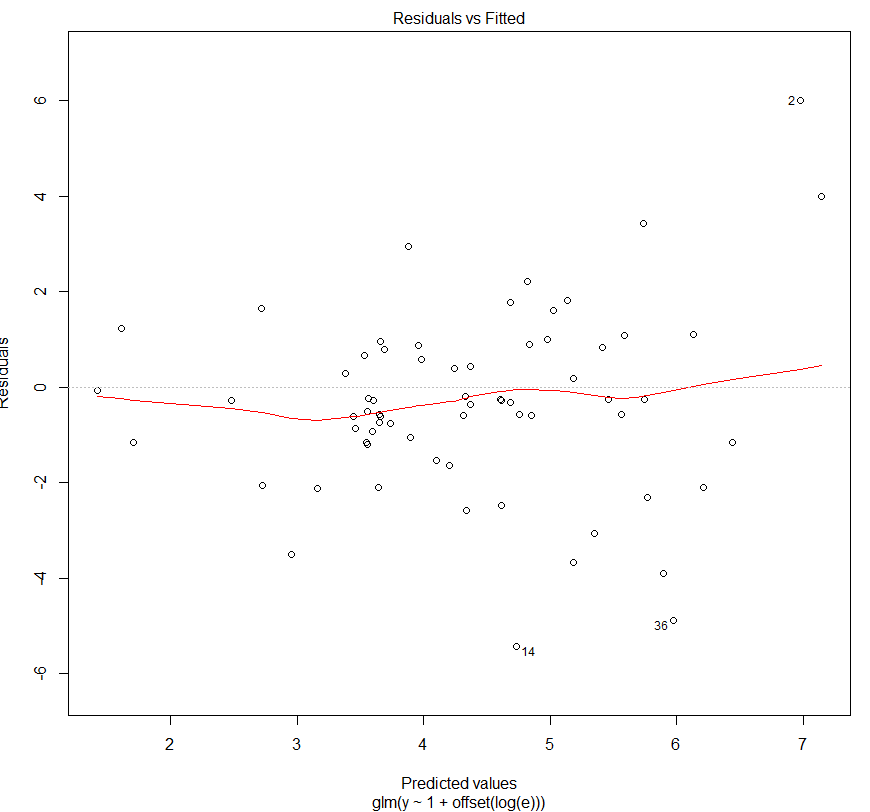
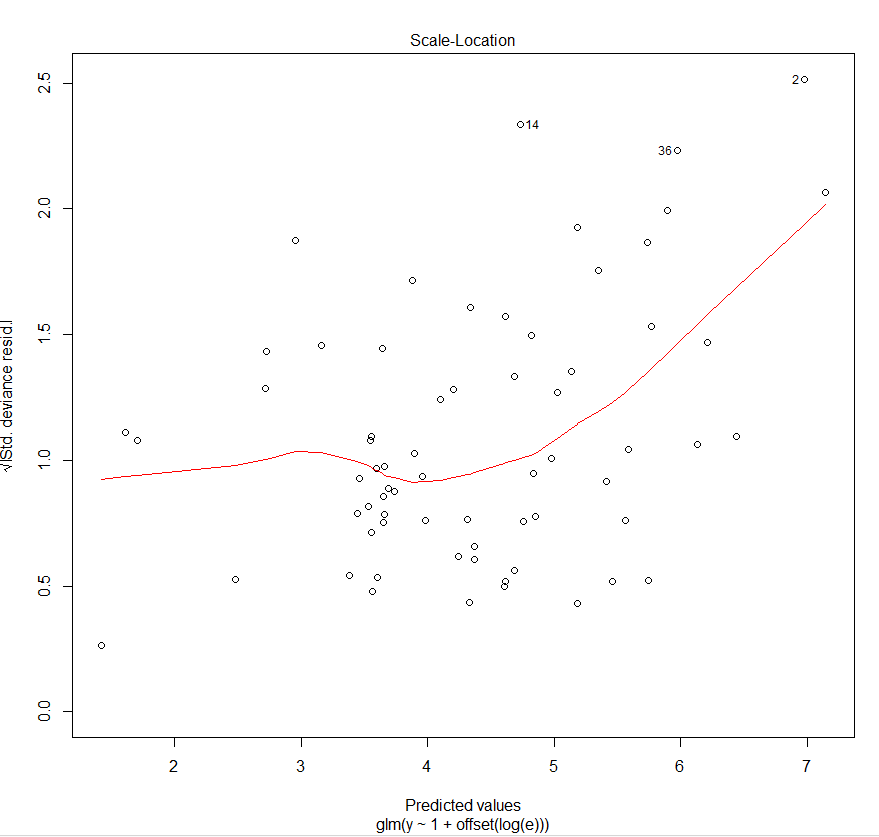
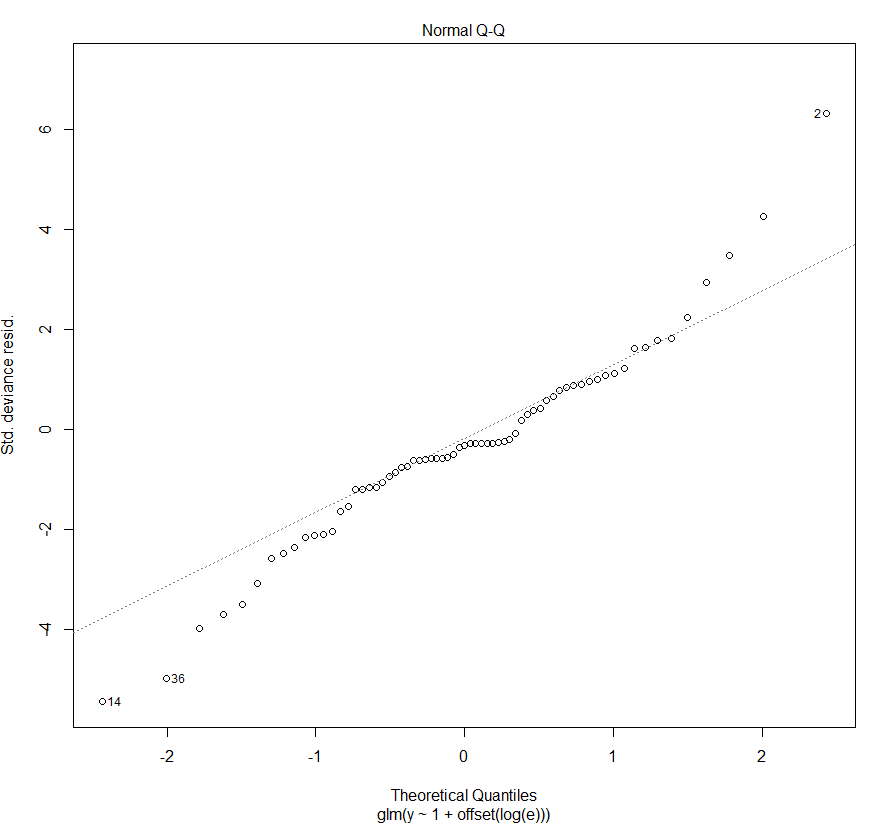
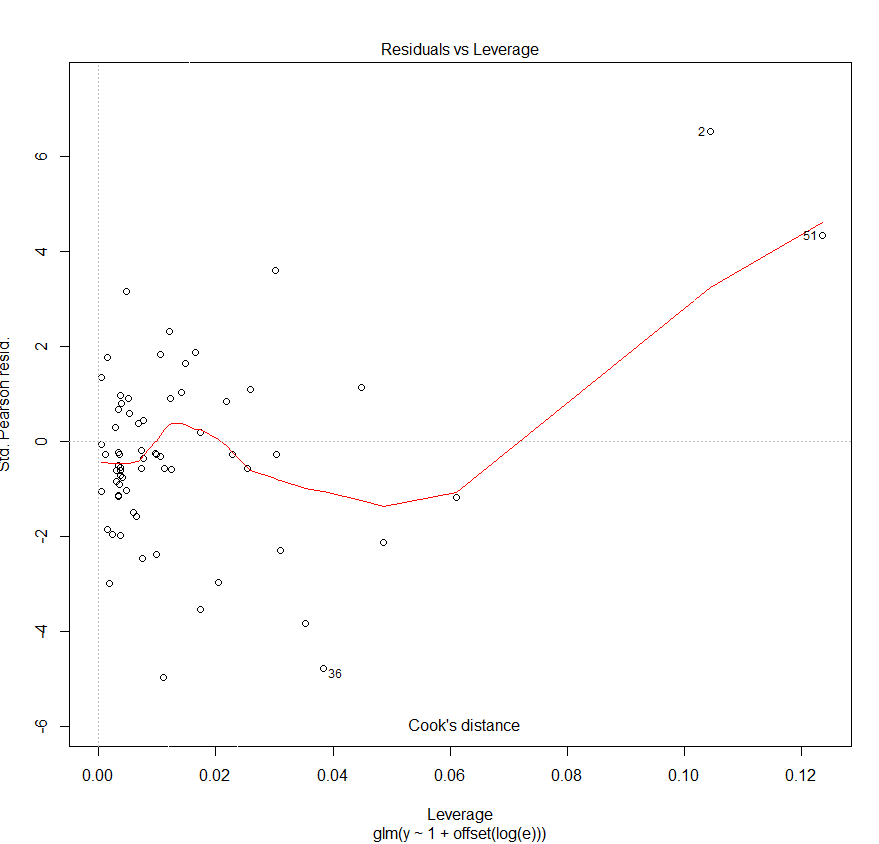


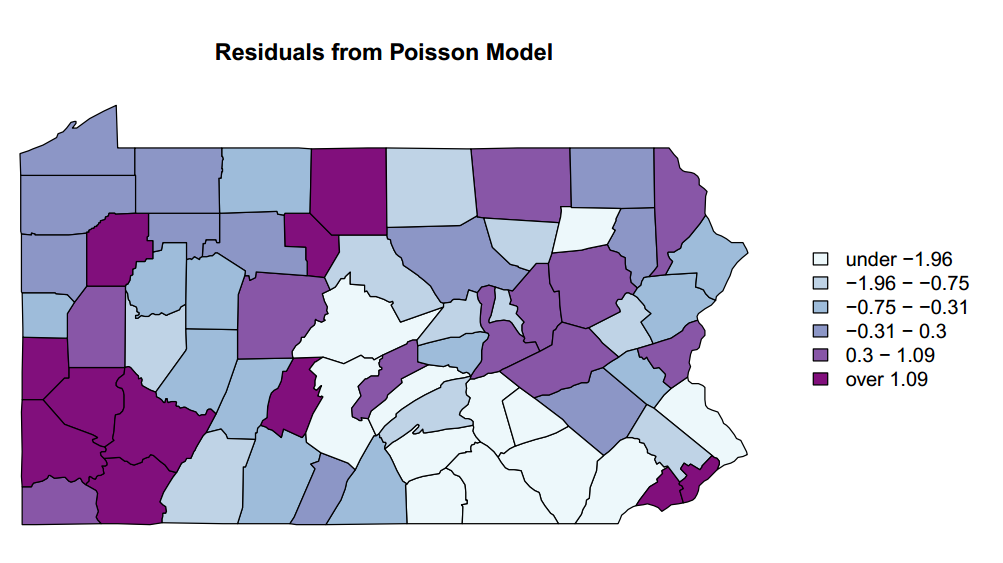
The estimates of SMR seem relatively stable with the exception of the northern counties. There are a few counties (highlighted in red) that have a SE > = 0.25. This instability is likely due to driven small population and sample sizes in those particular areas.

6.)

Estimate of Intercept : 2.19e-17

SE of Intercept : 9.86e-03



Just from an eyeball of the residuals mapped, there is likely clustering at the bottom left and right corners of Pennsylvania.

7.)

In this naïve model, with only an intercept term, the Moran’s I test across the various weighting schemes suggest that there is spatial autocorrelation in breast cancer incidence in Pennsylvania. The tests are not sensitive to different assumptions of the weighting scheme.

|  |  |  |
| --- | --- | --- |
| Weighting Scheme | Moran I Statistic | p-value |
| W | 0.223 | 0.0007 |
| B | 0.218 | 0.0005 |
| C | 0.219 | 0.0005 |
| U | 0.218 | 0.0005 |
| S | 0.221 | 0.0005 |

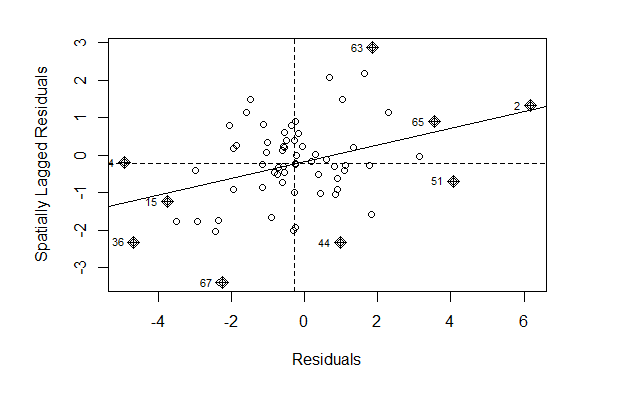
8.)

Geary’s C test across the differing weighting schemes similarly suggests that there exists spatial autocorrelation.

|  |  |  |
| --- | --- | --- |
| Weighting Scheme | Geary C Statistic | p-value |
| W | 0.767 | 0.0022 |
| B | 0.777 | 0.014 |
| C | 0.777 | 0.014 |
| U | 0.777 | 0.014 |
| S | 0.774 | 0.005 |

9.)

Below is a plot of the spatially lagged residuals versus the residuals. The plot highlights counties that are influential in the detected clustering with crosshairs.



The plot below maps local contributions that have been standardized and turned into a z-score. The map identifies counties—Washington and York County in particular, that are significantly influential to the detected clustering.

