IN SEARCH OF RISK FACTORS OF UTERINE SARCOMA

Phyon Christopher
Tenecia Harris
Dr. Brian Smith
University of Iowa
Department Of Biostatistics
TABLE OF CONTENTS

- Background information
- Data
- Objective
- Correlation
- Moran’s I
- Hypothesis Testing
- Simulations
- Cancer Data by Decades
- Conclusions and Future Work
- References
- Acknowledgements
BACKGROUND

A rare type of uterine cancer

+ It usually occurs after menopause.
+ Two main types:
  × leiomyosarcoma (cancer that begins in smooth muscle cells)
  × endometrial stromal sarcoma (cancer that begins in connective tissue cells)
+ Survival Rates: 89.5% 2-year, 84.7 5-year, 82.6 10-year
SURVEILLANCE, EPIDEMIOLOGY AND END RESULTS (SEER)  

9 REGISTRIES

- California
- Connecticut
- Atlanta (Metropolitan)
- Hawaii
- Iowa
- Detroit (Metropolitan)
- New Mexico
- Utah
- Washington
Our analysis goal is to test for evidence of spatial patterns in county uterine cancer rates, which could provide clues pointing to environmental risk factors for the disease.

Remove differences in county rates due to age and race in order to examine possible patterns due to environmental risk factors.

This is accomplished by first applying indirect standardization to the county rates.
INDIRECT STANDARDIZATION

- Compute the expected number of events
  \[ E = \sum \left( \frac{e_i^s}{p_i^s} \right) p_i = \sum r_i^s p_i \]

- Divide observed number of events by the expected number of events
  \[ SER = \frac{e}{E} \]

- The ISR is the product of the standardized event rate and the crude rate of the standard population.
  \[ ISR = SER \times r^s \]
Statistical correlation is a measure of the linear relationship between two random variables $X$ and $Y$.

Spatial autocorrelation is more complex because it is correlation as a function of proximity (and possibly direction) between observations on a single random variable $X$ measured in two-dimensions.
MORAN’S I

- Moran's I is one measure of spatial autocorrelation for a random variable X measured in two-dimensional space.
  - Function of proximity
- A Moran’s I close to 1 shows strong spatial autocorrelation; 0 no autocorrelation.
- Moran's I for this data is 0.071, suggestive of weak spatial autocorrelation.
HYPOTHESIS TESTING

H₀: Observations are spatially independent.

Hₐ: Observations are autocorrelated.

- Under the null, Moran’s I has an expected value and variance of:

\[ E(I) = -\frac{1}{N - 1} \]

\[ \text{var}(I) = \frac{N^2 S_1 - NS_2 + 3S_0^2}{(N - 1)(N + 1)S_0^2} - \left(\frac{1}{N - 1}\right)^2 \]
Assuming the data are normally distributed or the sample size is large, the null hypothesis can be evaluated with the test statistic

\[ z = \frac{I - E(I)}{\sqrt{\text{var}(I)}} \]

The distribution is not a normal distribution so we can use simulations to calculate p-values.
**SIMULATIONS**

- The general idea is to simulate the distribution of $I$ under the null hypothesis of no spatial autocorrelation; i.e. that county measurements are independent.
  1. Randomly reassign the measurements to counties.
  2. Compute the $I$ for each.
  3. Repeat steps 1 and 2 $K$ times.
  4. Calculate the $p$-values as the proportions of simulated $I$ that are larger than the values computed on the original data.
SIMULATED P-VALUE RESULT

Histogram

Frequency

Simulated I

0 500 1000 1500

-0.2 -0.1 0.0 0.1 0.2
The Moran's $I$ in previous slides tests for evidence of any spatial clustering:

- Does not indicate the location of clusters.

Local indicators of spatial association (LISAs) have been proposed to provide local measures of similarity between the value for a particular county and those neighboring it.
- Moran’s I is 0.029
- Simulated P-value is 0.235
- Moran’s $I$ is 0.023.
- Simulated P-value is 0.233.
- Moran’s $I$ is 0.159
- Simulated P-value is 0.005

The 90’s

ISR from 1990-1999 of Uterine Cancer In

Histogram of ISR

P-values for Clustering Around Counties

- $p < 0.05$
- $p < 0.01$
- Moran’s $I$ is 0.009
- Simulated P-value is 0.346
CONCLUSIONS AND FUTURE WORK

- In general, there was not much evidence of clustering.
- We would like to further investigate similarities in counties with significant evidence.
  - Adams, Clarke, and Wayne County all showed in at least one of the decades and the map for the overall number of years (1973-2006) on the simulated p-values plotted for clustering.
  - Median income and population density for these counties are lower than the median income and population density for the state.
    - Points to rural counties.
- Expand analysis to include county rates for all nine registries in the SEER registry.
REFERENCES


ACKNOWLEDGEMENTS

- Dr. Brian Smith
- University of Iowa Biostatistics Department
- The Alliance Program