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

IN SEARCH OF RISK FACTORS OF UTERINE SARCOMA

TABLE OF CONTENTS

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

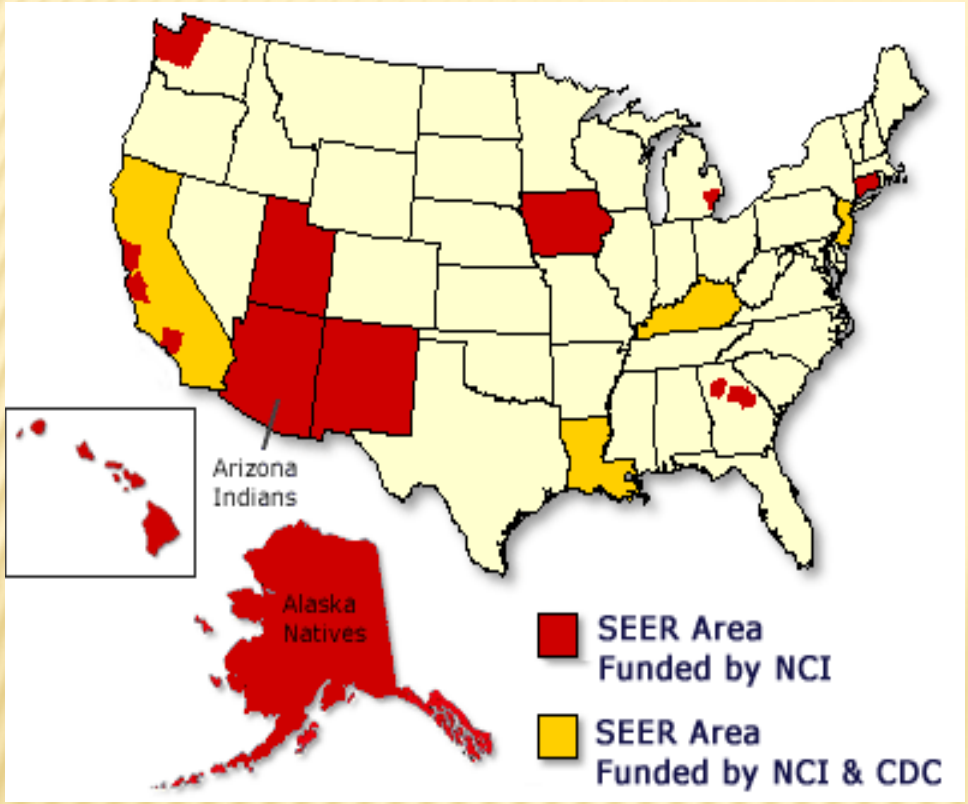
UTERINE SARCOMA

RATES PER 100,000

Age	Black	White	Other
00 years	0	0	0
01-04 years	0	0	0
05-09 years	0	0	0
10-14 years	0	0.013	0
15-19 years	0.077	0.038	0.072
20-24 years	0.088	0.041	0.136
25-29 years	0.336	0.166	0.217
30-34 years	0.793	0.406	0.398
35-39 years	1.361	0.858	0.737
40-44 years	2.367	1.508	1.469
45-49 years	2.367	2.571	2.657
50-54 years	3.578	3.033	2.809
55-59 years	6.944	3.534	2.815
60-64 years	10.465	4.549	3.327
65-69 years	11.934	5.478	4.613
70-74 years	15.884	5.478	4.876
75-79 years	12.87	6.496	4.717
80-84 years	14.175	6.816	3.087
85+ years	9.652	5.046	1.455

DATA

SURVEILLANCE, EPIDEMIOLOGY AND END RESULTS (SEER)



9 REGISTRIES

- California
- Connecticut
- Atlanta (Metropolitan)
- Hawaii
- Iowa
- Detroit (Metropolitan)
- New Mexico
- Utah
- Washington

OBJECTIVE

- 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

$$E = \sum \left(\frac{e_i^s}{p_i^s} \right) p_i = \sum r_i^s p_i$$

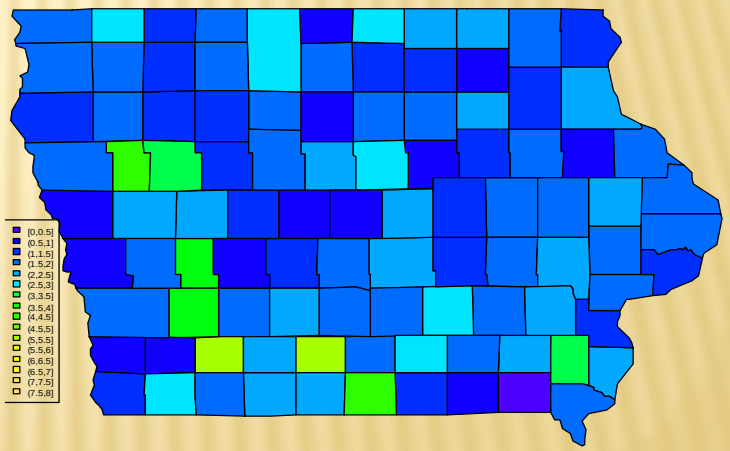
- Divide observed number of events by the expected number of events

$$SER = \frac{e}{E}$$

- The ISR is the product of the standardized event rate and the crude rate of the standard population.

$$ISR = SER \times r^s$$

ISR of Uterine Cancer Incidence in Io



DATA

CORRELATION

- 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

H_0: Observations are spatially independent.

H_A: Observations are autocorrelated.

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

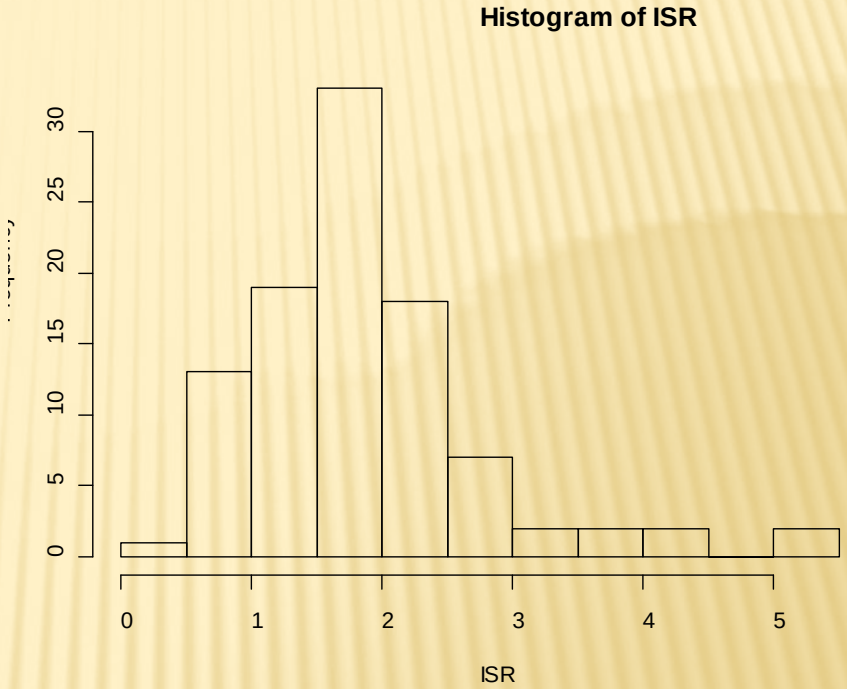
$$E(I) = -\frac{1}{N-1}$$

$$\text{var}(I) = \frac{N^2 S_1 - N S_2 + 3 S_0^2}{(N-1)(N+1) S_0^2} - \left(\frac{1}{N-1}\right)^2$$

HYPOTHESIS TESTING

- Assuming the data are normally distributed or the sample size is large, the null hypothesis can be evaluated with the test statistic

$$z = \frac{I - \mathrm{E}(I)}{\sqrt{\mathrm{var}(I)}}$$



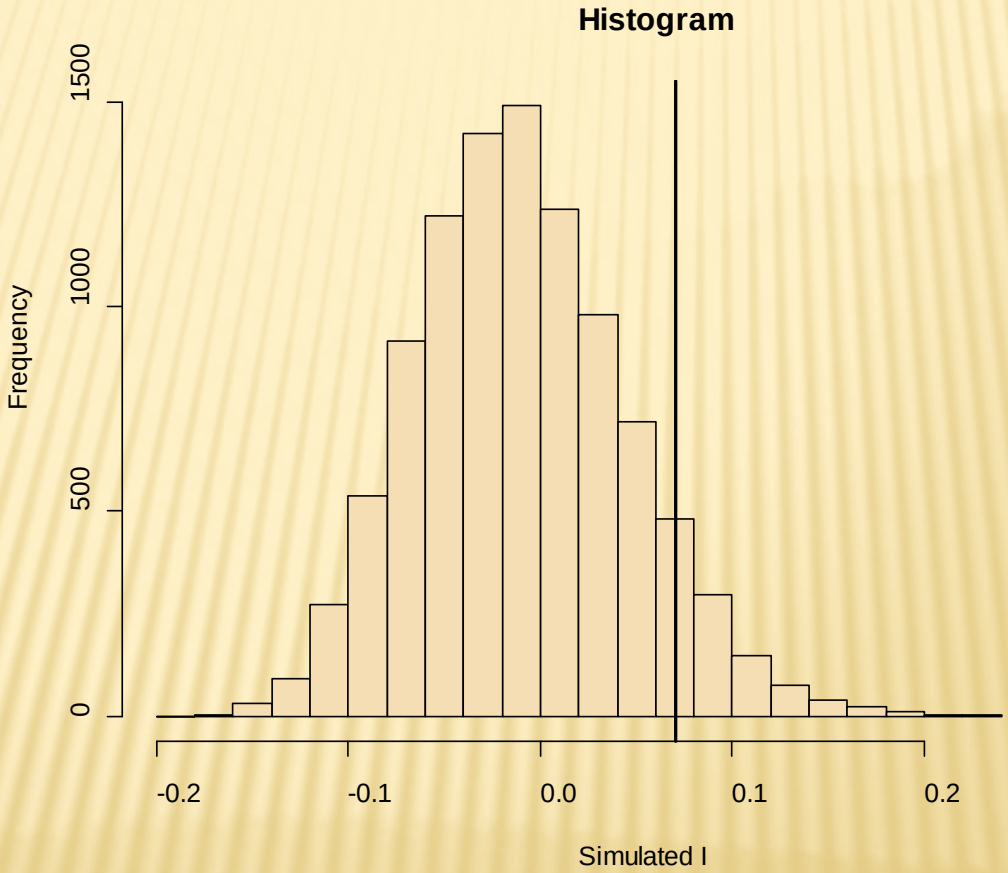
- The distribution is not a normal distribution so we can use simulations to calculate p-values.

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.

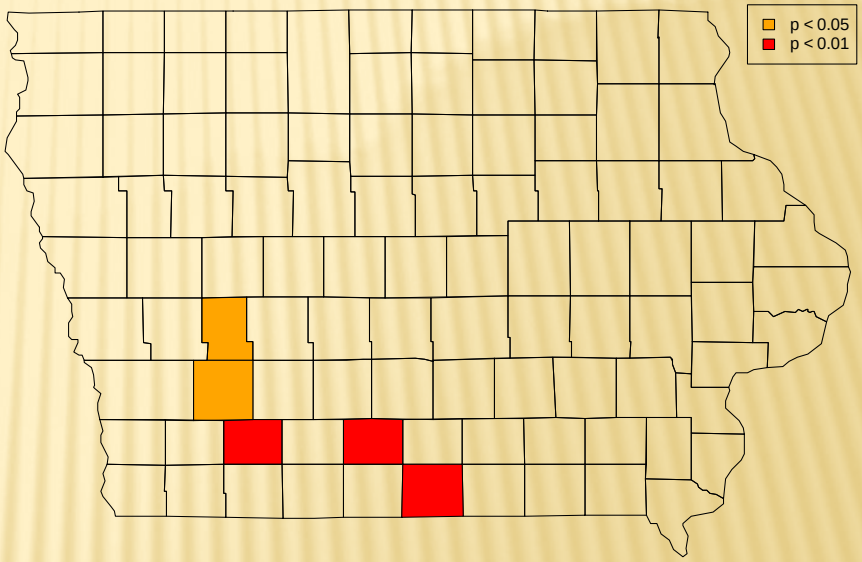
SIMULATED P-VALUE RESULT



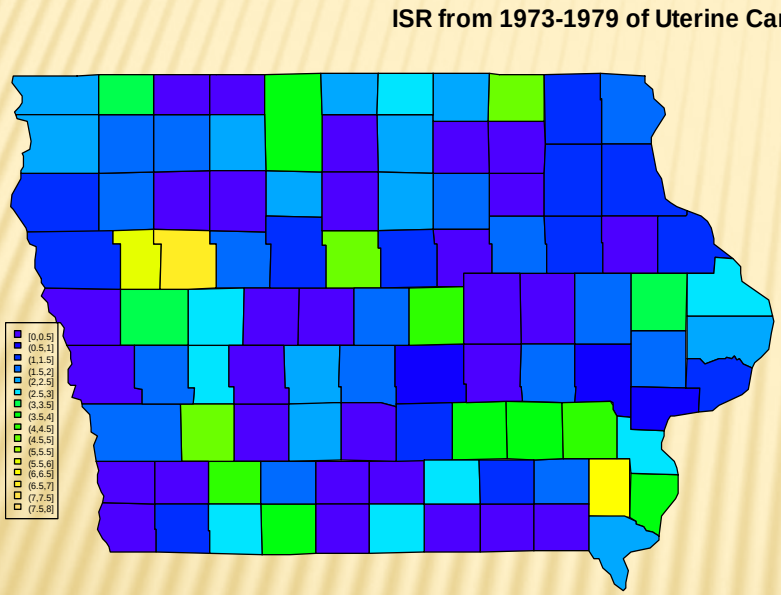
SIMULATIONS

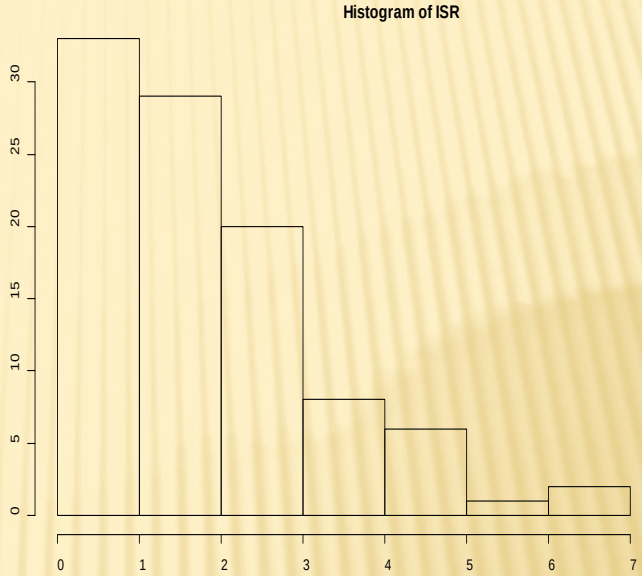
- The Moran's *I* in previous slides tests for evidence of any spatial clustering
  - Does not indicate the location of clusters.
- Local indicators of spatial association (LISAs) have been proposed to provide local measures of similarity between the value for a particular county and those neighboring it.

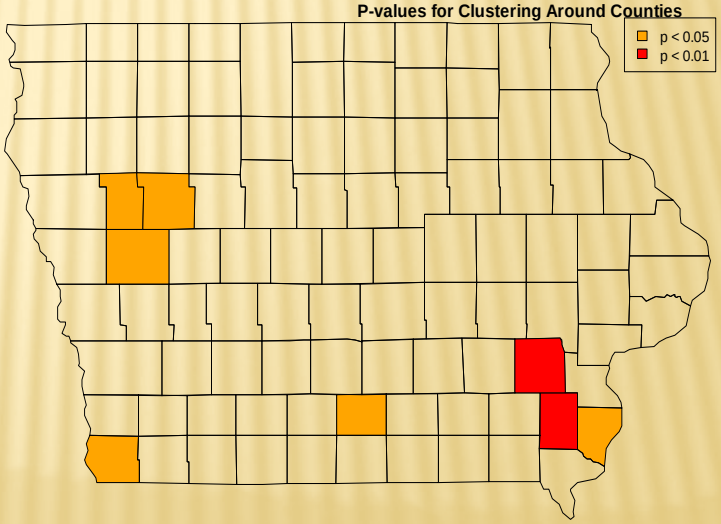
P-VALUES OF LOCAL MORAN'S I



- Moran's *I* is 0.029
- Simulated P-value is 0.235



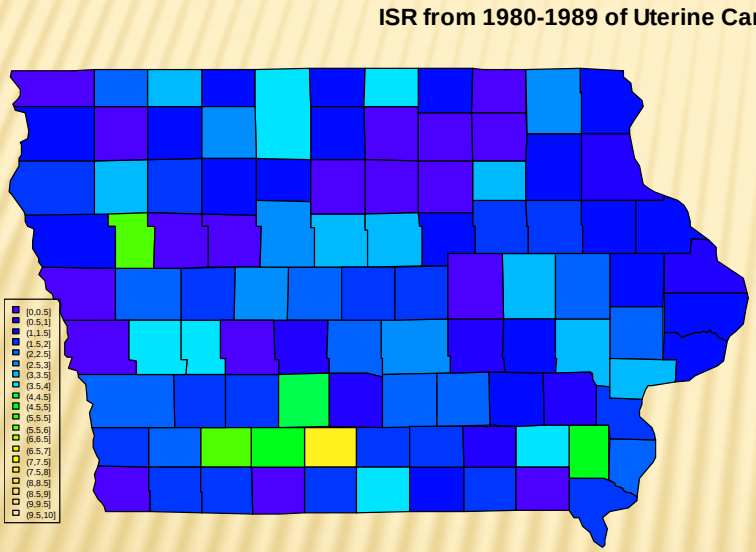


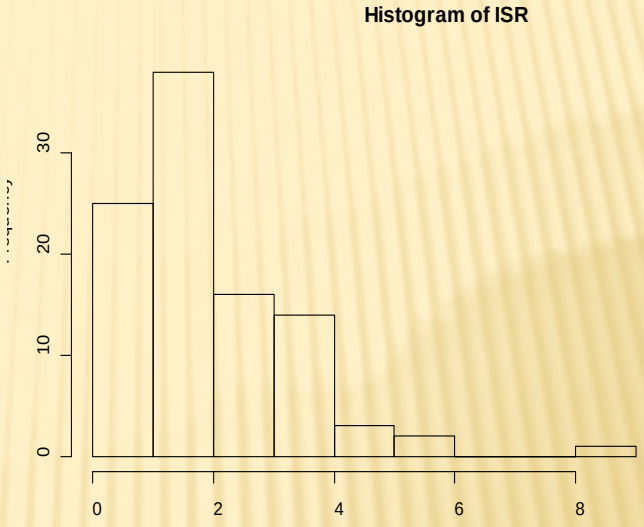


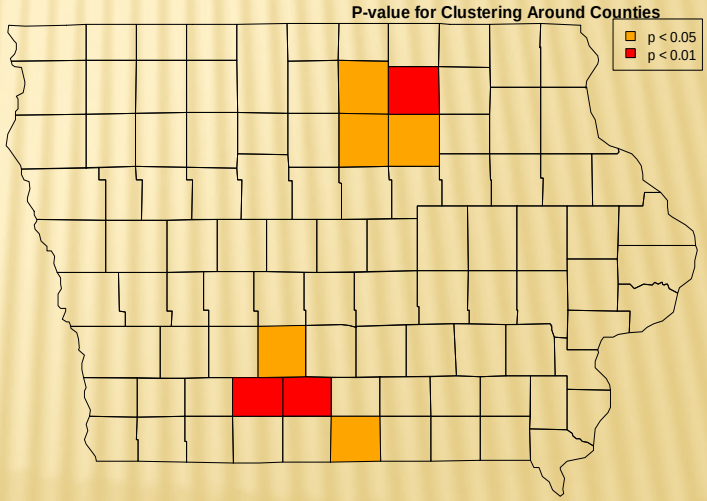
1973-1979

THE 80’S

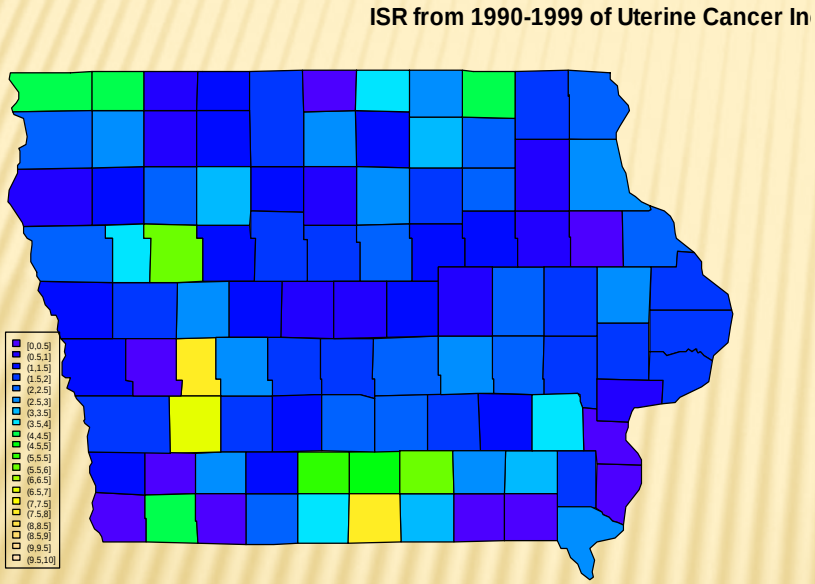
- Moran’s *I* is 0.023.
- Simulated P-value is 0.233.

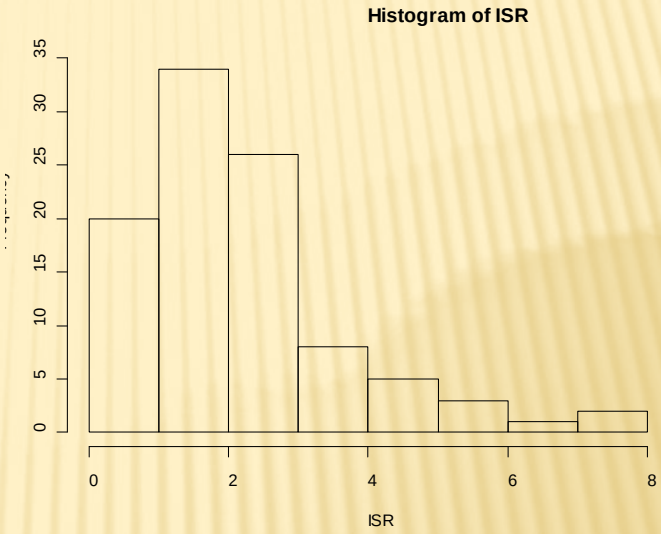


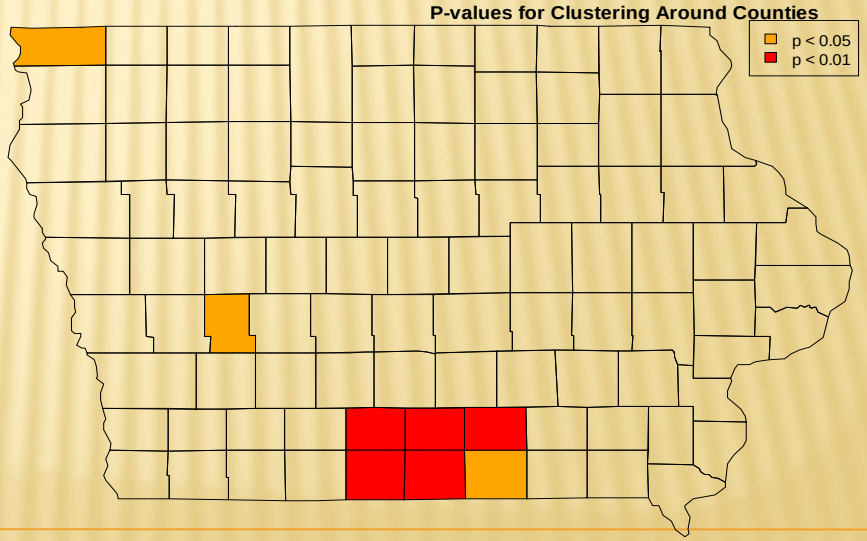




- Moran's *I* is 0.159
- Simulated P-value is 0.005

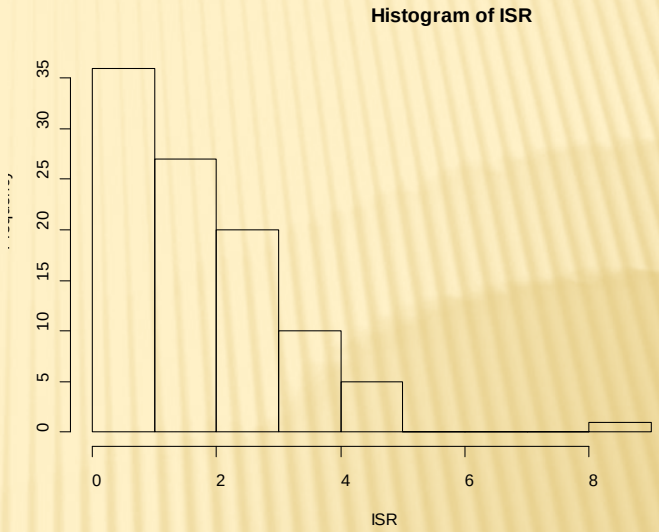


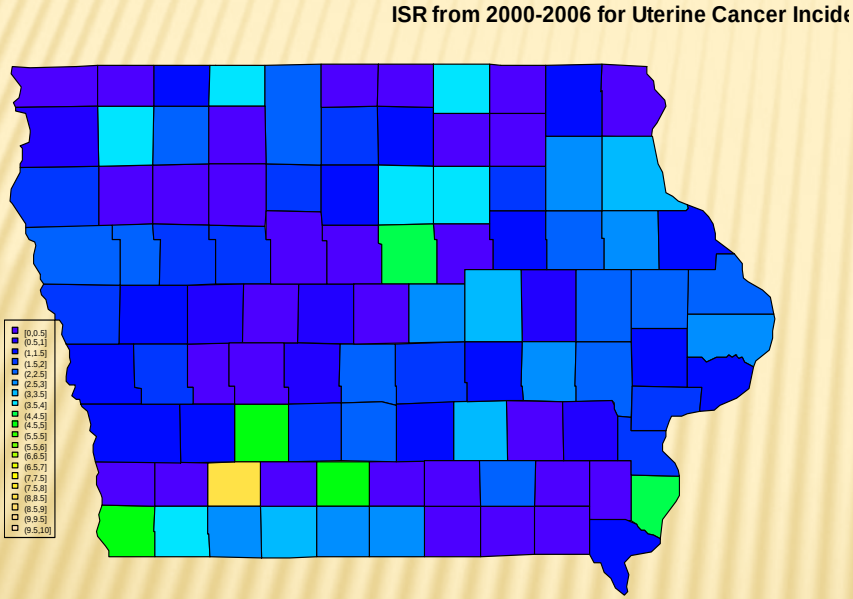


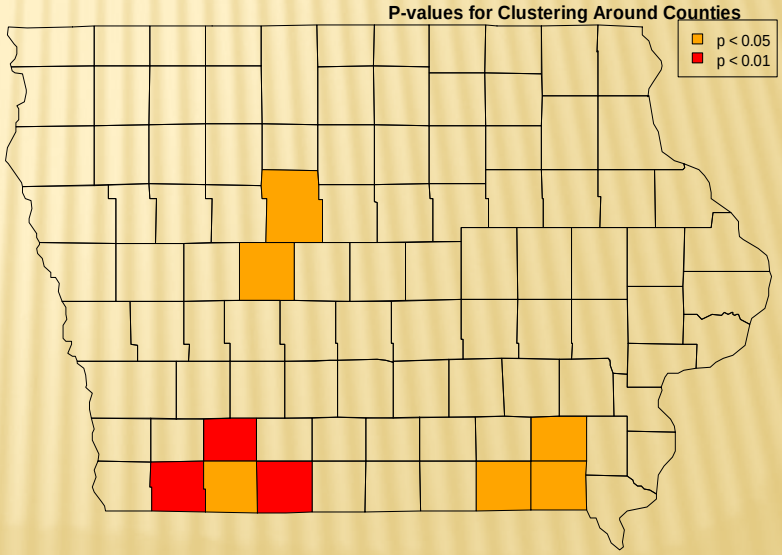


THE 90'S

- Moran's *I* is 0.009
- Simulated P-value is 0.346







2000-2006

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 a 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

- Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2008 Sub (1973-2006) Total U.S., 1973-2006 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009, based on the November 2008 submission.
- Ries LAG, Young JL, Keel GE, Eisner MP, Lin YD, Horner M-J (editors). SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics. National Cancer Institute, SEER Program, NIH Pub. No. 07-6215, Bethesda, MD, 2007.
- Waller, A.L. and Gotway, C.A. (2004) Applied Spatial Statistics for Public Health Data, New Jersey: John Wiley & Sons, Inc.

ACKNOWLEDGEMENTS

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