Spatial Data Analysis

Moran's I Tests for Global Autocorrelation

Sergio Rey

http://cogs.sdsu.edu/

Outline

- Moran's I
- Moran Scatter Plot
- Geary's c

Moran's I

Moran's I

$$I = \left(\frac{n}{S_0}\right) \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

- $z_i = x_i - \bar{x}$
- Cross-product statistic
- Similar to a correlation coefficient
- Value depends on W

Moran's I

- Scaling Factors
  - denominator: $n =$ number of observations
  - numerator: $S_0 = \sum_i \sum_j W_{ij}$
- S_0
  - number of nonzero elements in W
  - number of neighbor pairs $(\times 2)$

Inference

- Analytical
  - exact
  - approximate: normal and equal probability (randomization)
- Computational
  - random permutations

Exact Inference

- Assume Normality
  - for variable
- Ratio of Normal Quadratic Forms
  - Tiefelsdorf-Boots

$$I = y'Wy/y'y$$

  - Exact distribution depends on eigenvalues of W

Analytical Inference

- Normal
  - assume uncorrelated normal distribution
- Randomization
  - each observation equally likely to fall on each location
- Standardization
  - Compute: $E[I]$ and $V[I]$
- $z = (I - E[I])/\sqrt{V[I]}$

Normal Approximation

- Mean

$$E[I] = -1/(n-1)$$

  - **Not** zero, but approaches zero as $n \to \infty$
  - does not depend on W or y

Second Moment

$$V[I] = E[I^2] - E[I]^2$$

$$E[I_N^2] = n^2 S_1 - n S_2 + 3S_0^2 / \left[S_0^2 \left(n^2 - 1\right)\right]$$

$$S_1 = (1/2) \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{i,j} + w_{j,i})^2$$

$$S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{i,j} + \sum_{j=1}^{n} w_{j,i} \right)^2$$

Second Moment

$$V[I] = E[I^2] - E[I]^2$$

$$V(I_N) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n-1)(n+1) S_0^2} - (1/(n-1))^2$$

Solely a function of W

Randomization

- Mean

$$E[I] = -1/(n-1)$$

  - same as for normal approximation

Second Moment

$$E[I_R^2] = \frac{n\left[\left(n^2 - 3n + 3\right)S_1 - nS_2 + 3S_0^2\right] - b_2\left[\left(n^2 - n\right)S_1 - 2nS_2 + 6S_0^2\right]}{(n-1)(n-2)(n-3)S_0^2}$$

$$b_2 = \frac{\sum_{i=1}^{n} z_i^4}{\left(\sum_{i=1}^{n} z_i^2\right)^2}$$

- depends on weights and distribution of variable

Computational Inference

- Permutation Approach
  - reshuffle observations
  - construct reference distribution from random permutations
  - pseudo significance

$$p = \frac{M+1}{n+1}$$

Example n=16
Continuous y

39	41	38	45
39	42	41	42
48	49	48	51
47	51	50	55

Example: Moran's I

i	y_i	$z_i = y_i - \bar{y}$	z_i^2	z_i^4	$\sum_j w_{ij} z_i z_j$
1	39	-6.375	40.640625	1.651660e+03	68.531250
2	41	-4.375	19.140625	3.663635e+02	74.921875
3	38	-7.375	54.390625	2.958340e+03	67.296875
4	45	-0.375	0.140625	1.977539e-02	4.031250
5	39	-6.375	40.640625	1.651660e+03	45.421875
6	42	-3.375	11.390625	1.297463e+02	38.812500
7	41	-4.375	19.140625	3.663635e+02	50.312500
8	42	-3.375	11.390625	1.297463e+02	-2.953125
9	48	2.625	6.890625	4.748071e+01	-2.953125
10	49	3.625	13.140625	1.726760e+02	27.187500
11	48	2.625	6.890625	4.748071e+01	24.937500
12	51	5.625	31.640625	1.001129e+03	49.921875
13	47	1.625	2.640625	6.972900e+00	13.406250
14	51	5.625	31.640625	1.001129e+03	55.546875
15	50	4.625	21.390625	4.575588e+02	82.671875
16	55	9.625	92.640625	8.582285e+03	98.656250
$\sum_i$	726	0	403.75	18570.61	695.75
				$\bar{y}$	45.375
				$\sum_i \sum_j w_{i,j}$	48

Example: Moran's I

Component	Value
n	16
S_0	48
S_1	96
S_2	608
$E(I)$	-0.06667
$V(I_N)$	0.0326
b_2	0.1139203
$E(I_R^2)$	0.04404776
$V(I_R)$	0.0396

Moran's I

$$I = \left(\frac{n}{S_0}\right) \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

$$I = \left(\frac{16}{48}\right) \frac{695.75}{403.75} = 0.5744066$$

Example

$$
\begin{aligned}
E(I) &= \frac{-1}{16-1} = -0.06667 \\
V(I_N) &= \frac{16^2(96) - 16(608) + 3(48)^2}{(16-1)(16+1)48^2} - (1/(16-1))^2 = 0.03259 \\
z(I_N) &= \frac{I - E(I)}{\sqrt{V(I_N)}} = \frac{0.5744066 + 0.0667}{\sqrt{0.03259}} = 3.568112 \\
z(I_R) &= \frac{I - E(I)}{\sqrt{V(I_R)}} = \frac{0.5744066 + 0.0667}{\sqrt{0.03960331}} = 3.221380
\end{aligned}
$$

Standardized Statistic

- Moran's I Value Depends on W
  - value as such not comparable across tests
  - Use z-values
    - moments: E[I] and SE[I]
    - z_i = (I - E[I]) / SE[I]
    - Note: E[I] = -1 / (n - 1) such that mean under the null is NOT zero in small samples

Interpretation of Moran's I

- For Significant Statistics Only
- Use z-value
  - I depends on W
- Positive S.A. $z_i > 0$ for $p < 0.05$ ....
  - no distinction between clustering of high or low values
- Negative S.A. $z_i < 0$ for $p < 0.05$ ...

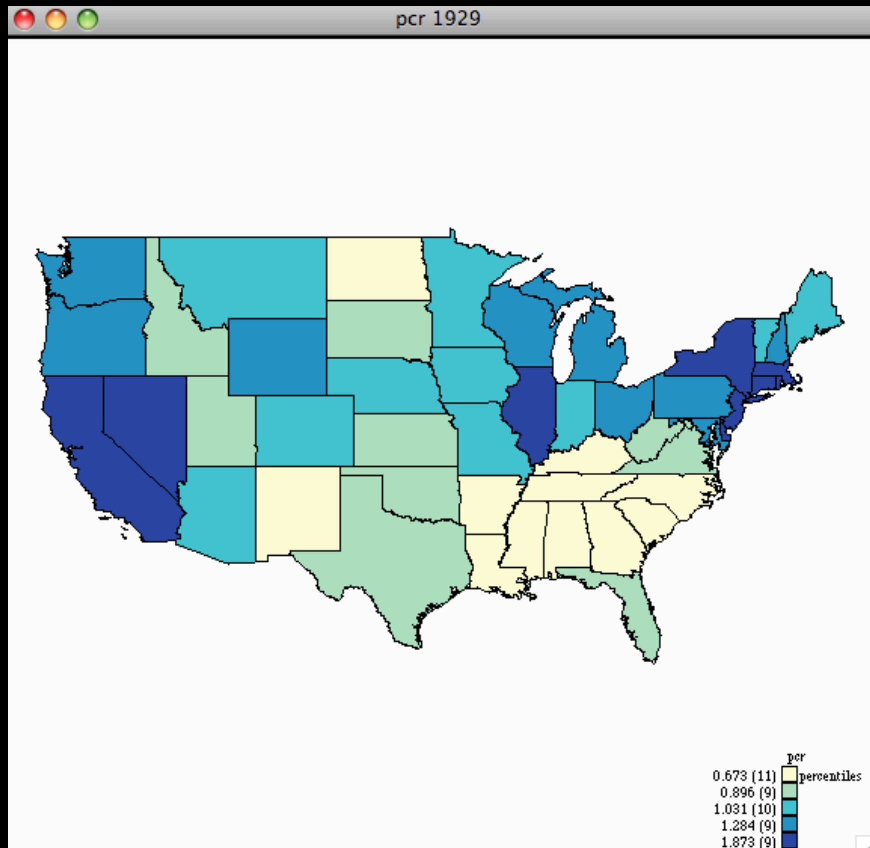
Moran's I for Rates

- Variance Instability of Rates: $p=x/n$
  - non-constant variance violates assumption of stationarity
  - may lead to spurious indication of spatial autocorrelation
- Empirical Bayes Adjustment
  - borrowing of information

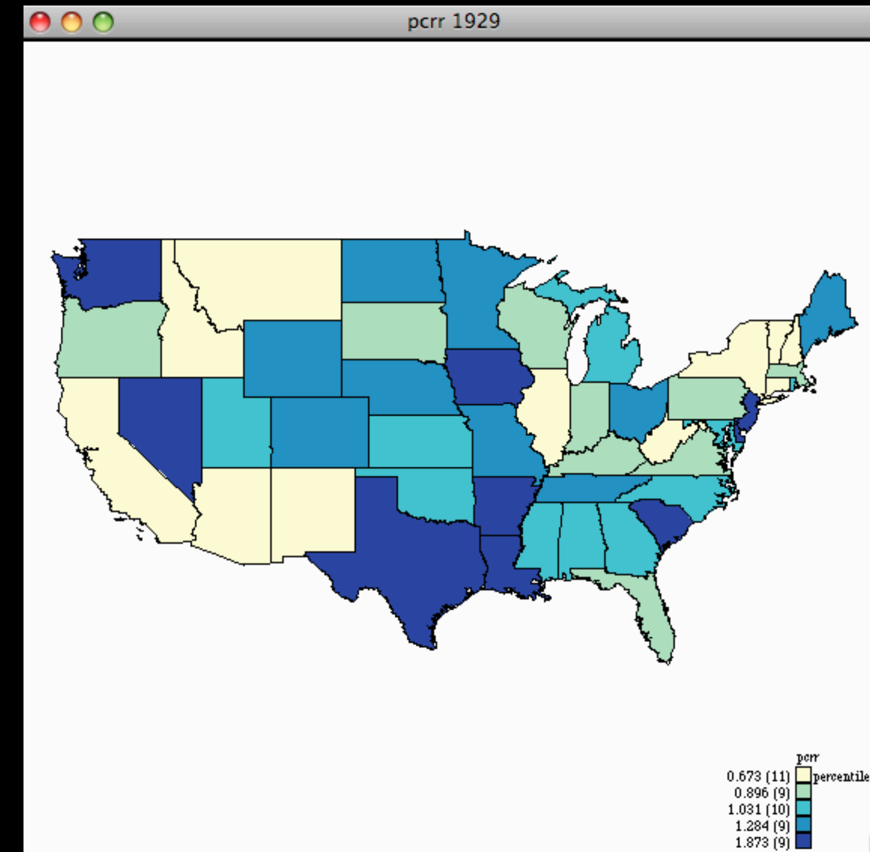
Moran Scatter Plot

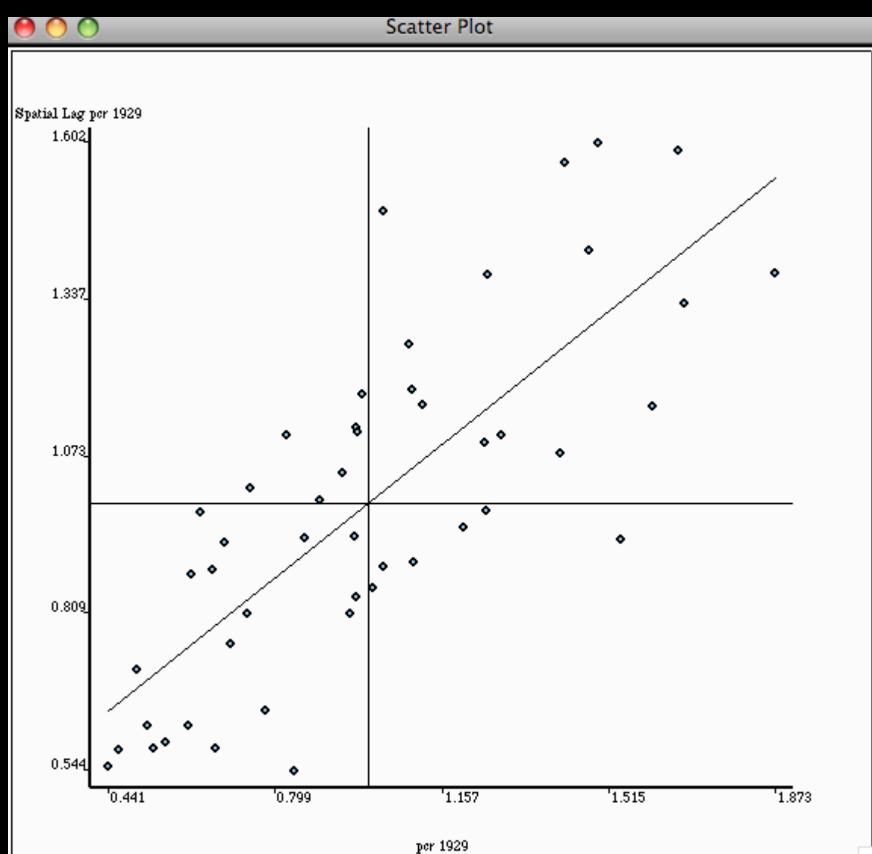
Principle

- Moran's I as a Regression Slope
  - in matrix notation: $I = z'Wz / z'z$
  - regression slope: $Wz = a + I.z$
- Moran Scatter Plot
  - linear association between Wz on axis and z on the axis
  - each point is pair (z_i, Wz_i), slope is I

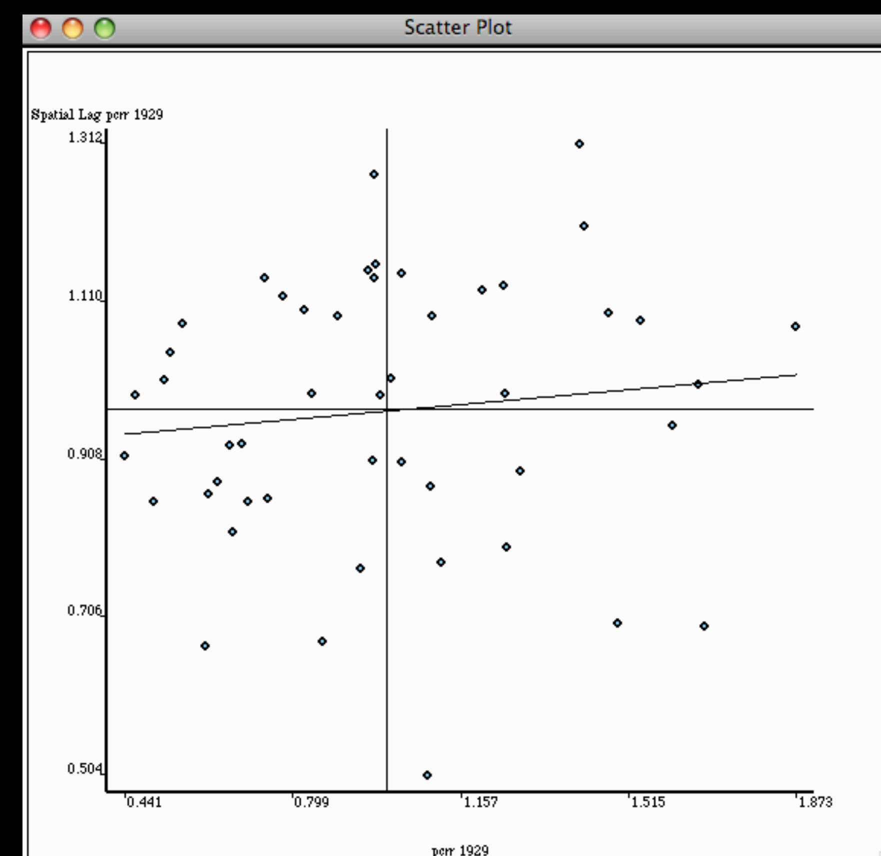


Locationally
Variant



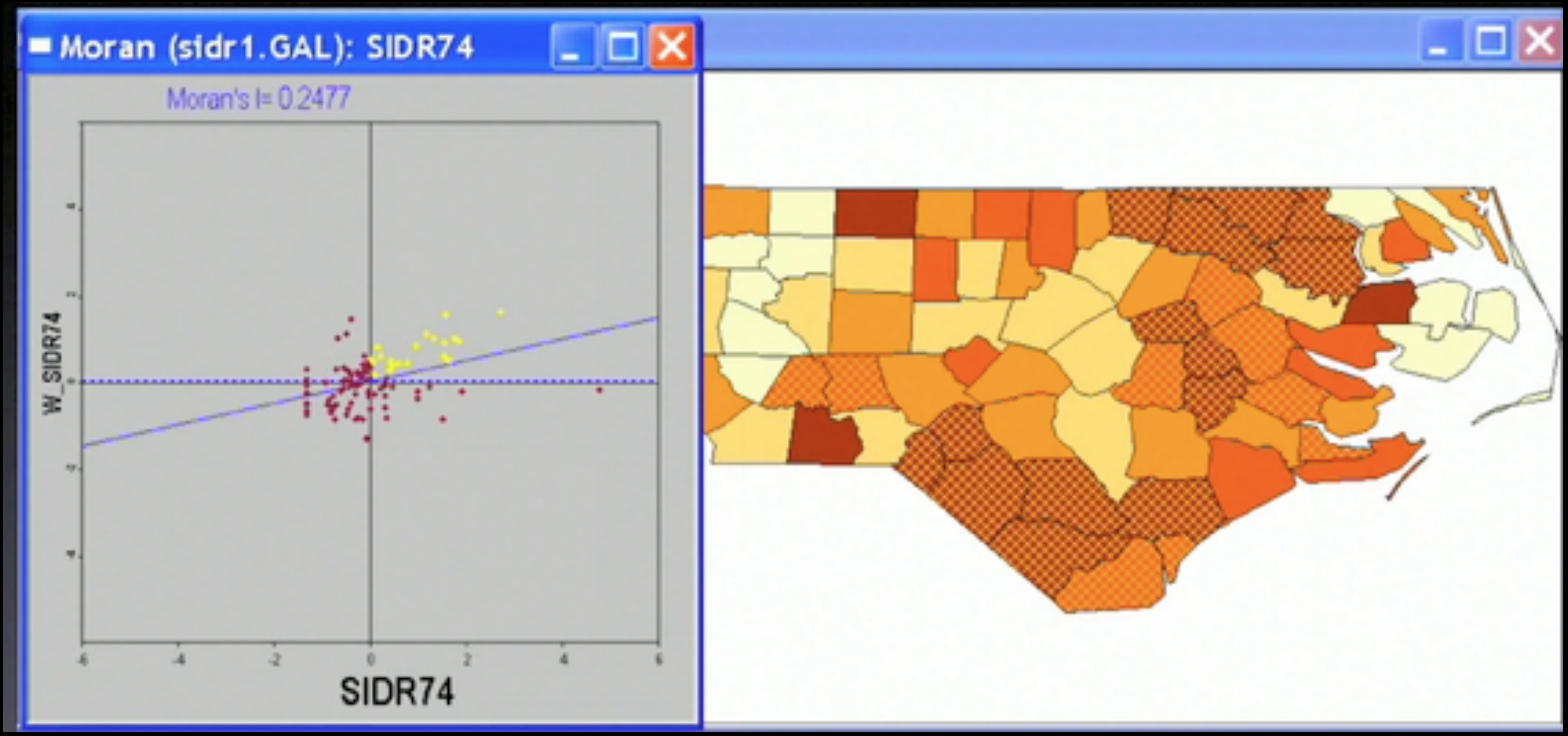


Locationally
Variant

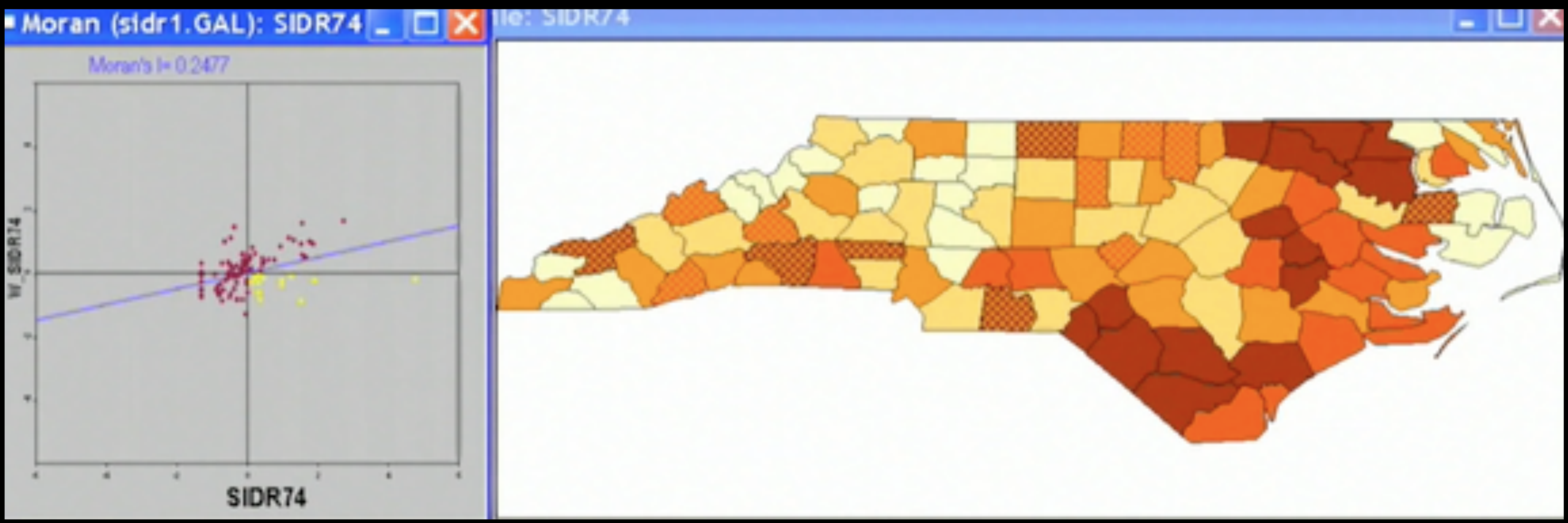


Link with Local SA

- Four Categories of SA
- Positive Spatial Autocorrelation
  - high-high and low-low: spatial clusters
- Negative Spatial Autocorrelation
  - high-low and low-high: spatial outliers
- Only Suggestive
  - no inference (yet)
  - relative to the mean



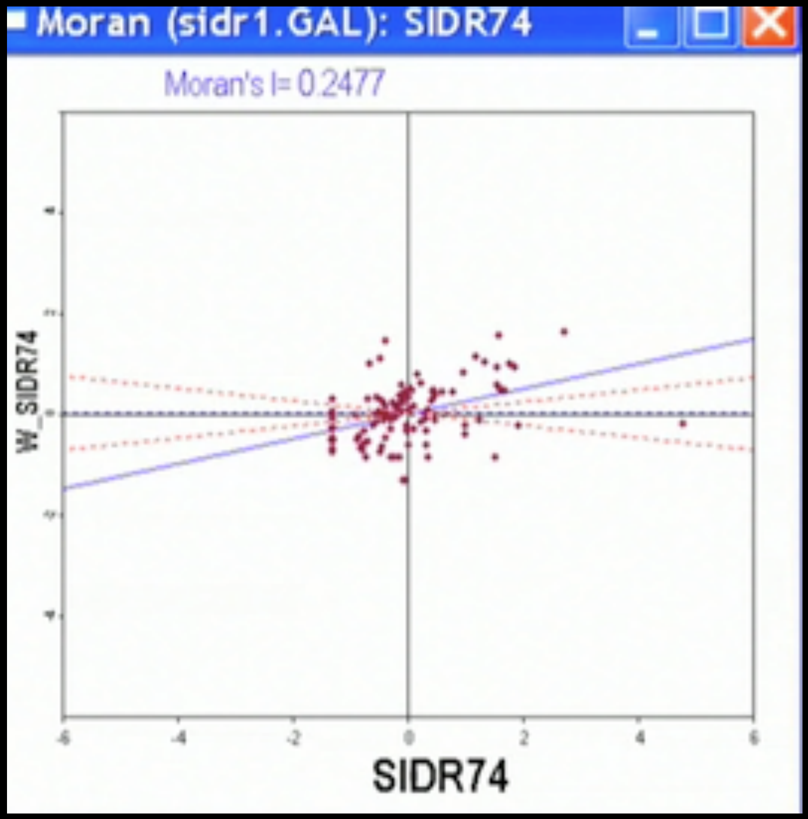
High-High Quadrant: Spatial Clusters

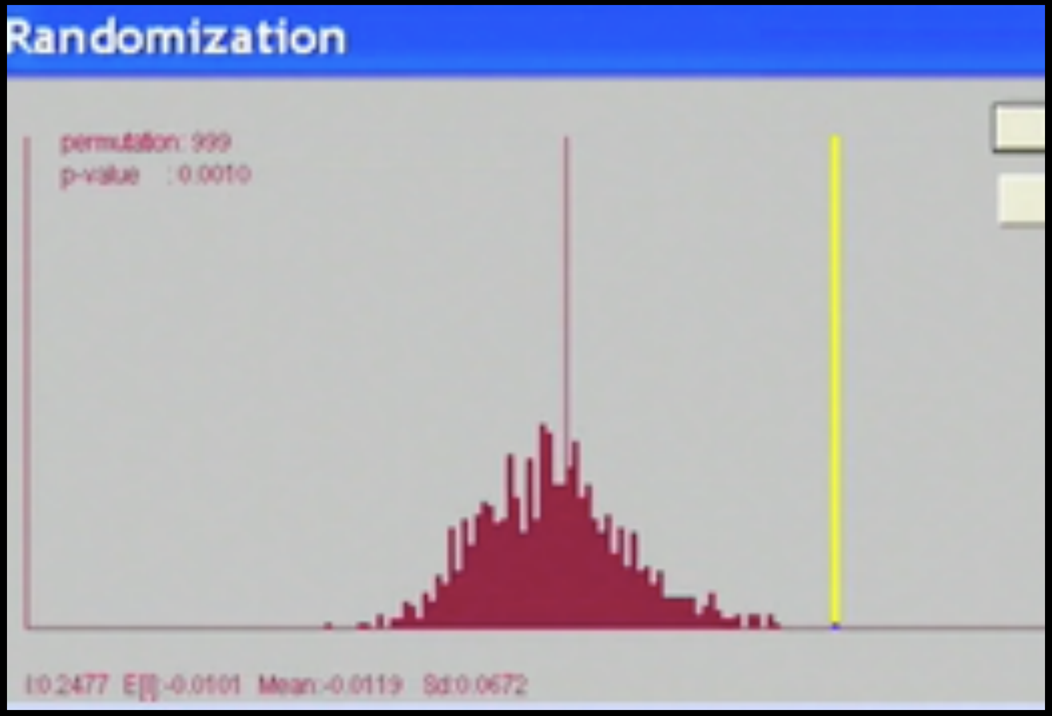


High-Low Quadrant: Spatial Outliers

Other Use of Moran Scatterplot

- Detect Local Nonstationarity
  - outliers in the scatter plot
  - high leverage points
  - sensitivity to boundary values
- Brushing the Moran Scatter Plot
  - different slopes in subsets of the data
  - suggest spatial regimes





Randomization Envelopes

Geary’s c

Geary's c

- Squared Difference
  - dissimilarity
  - similar to variogram
  - values between 0 and 2
- Statistic
- $c=(n-1) \sum\sum w_{ij} (x_i - x_j)^2 / 2 S_0 \sum z_i^2$

Inference

- Convert Geary's c to a z-value
  - distribution of statistic under the null of spatial randomness
  - $z = (c - E[c]/SE[c])$
- Moments of Geary's c
  - analytical: normal, randomization
  - computational: permutation

Interpretation of c

- For Significant Statistics Only
  - use z
- Positive Spatial Autocorrelation
  - $c < 1$ or $z < 0$
  - spatial clustering
  - opposite sign of Moran's I
- Negative Spatial Autocorrelation
  - $c > 0$ or $z > 0$
  - checkerboard pattern, competition

Moments: Normal Approximation

- Mean
  - $E[c] = 1$
  - does not depend on n, W, or y
- Variance
  - $Var[c] = [(2S_1 + S_2)(n-1) - 4 S_0^2] / [2(n+1)S_0^2]$
  - does not depend on y
  - only W and sample size

Moments: Randomization Approximation

- Equal Probability
- Mean
  - $E[c] = 1$
  - does not depend on W, y, or sample size
  - Variance
  - $V[c]=\{(n-1)\ S_1[n^2-3n+3-(n-1)b_2]$

    $-\ (1/4)(n-1)\ S_2[n^2+3n-6-(n^2-n+2)b_2]$

    $+(S_0^2[n^2-3-(n-1)2b_2]\}/[n(n-2)(n-3)S_0^2]$