Center for Spatially Integrated Social Science

Spatial Autocorrelation (3)
Global Spatial Autocorrelation

Luc Anselin
Spatial Analysis Laboratory
Dept. Agricultural and Consumer Economics
University of Illinois, Urbana-Champaign
http://sal.agecon.uiuc.edu

Outline

- Principles
- Join Count Statistic
- Moran's I
- Moran Scatterplot
- Bivariate/Space-Time Extension
- Geary's c

Principles

General Cross-Product Statistic

- Combinatorial Data Analysis
  - N objects $S = \{ o_1, o_2 ,\dots, o_N \}$
  - measures of "similarity" between objects
    - correlation matrix, distance matrix
  - are indications significantly different ?
  - association between two proximity matrices $A = \{ a_{ij} \}$ and $B = \{ b_{ij} \}$
  - gamma statistic $\Gamma_{AB} = \Sigma_i \Sigma_j a_{ij}b_{ij}$

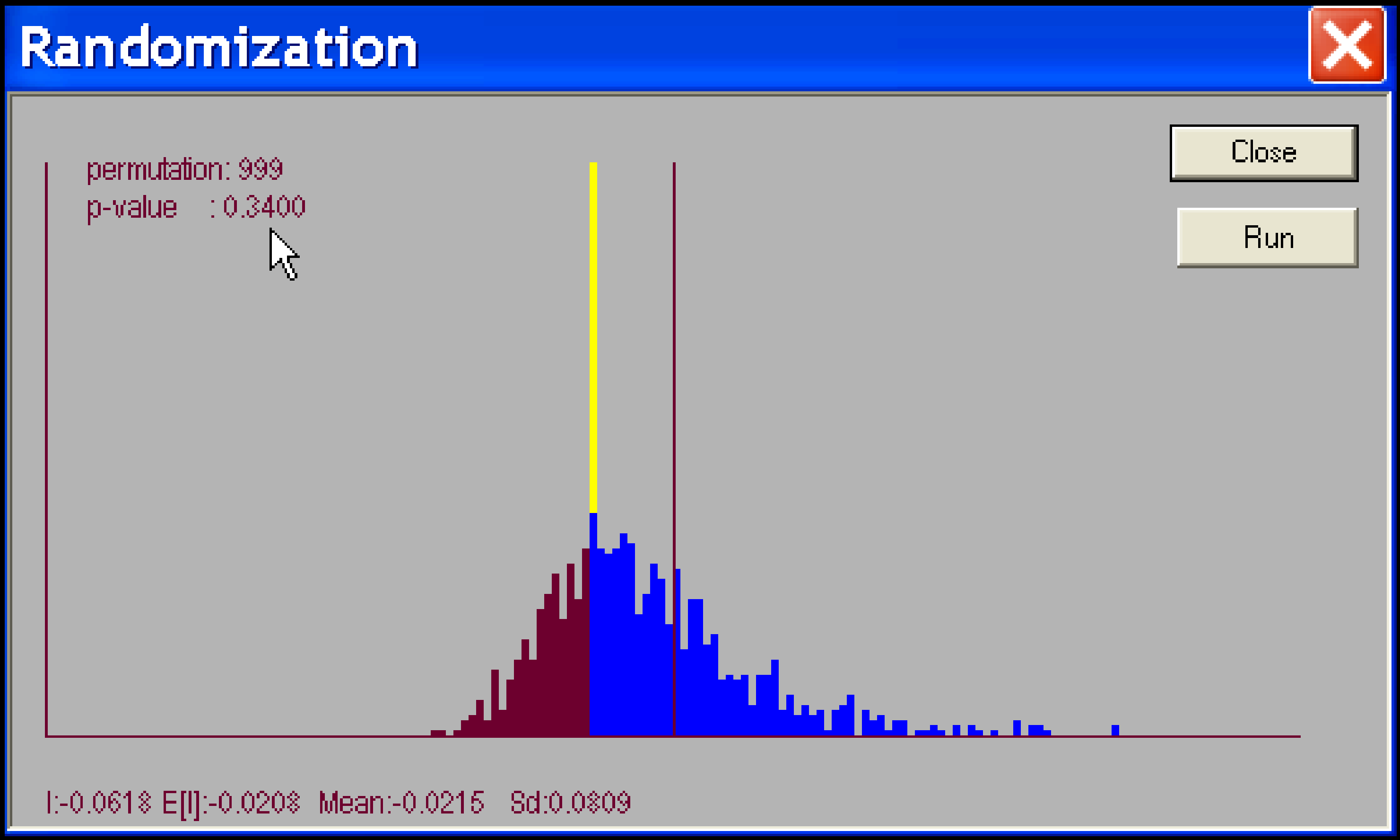
Gamma Index of Spatial Autocorrelation

- Match Between Locational Similarity and Value Similarity
  - locational similarity: W
  - value similarity: A
  - $\Gamma = \Sigma_i \Sigma_j w_{ij}.a_{ij}$
- Types of Value Association
  - cross-product: $x_i.x_j$
  - squared difference: $(x_i - x_j)^2$
  - absolute difference: $| x_i - x_j |$

Randomization Strategy

- Empirical Distribution Function
  - permute arrangement of objects
    - associate values with locations
    - associate locations with values
  - recomputed A and Γ_{AW}
  - compare observed Γ to distribution of Γ_{AW}
- Pseudo-Significance
  - $p = (T + 1) / (M + 1)$
    - M: # permutations; T: # times $\Gamma_{AW} \geq \Gamma$

Reference Distribution



Pros and Cons of Randomization

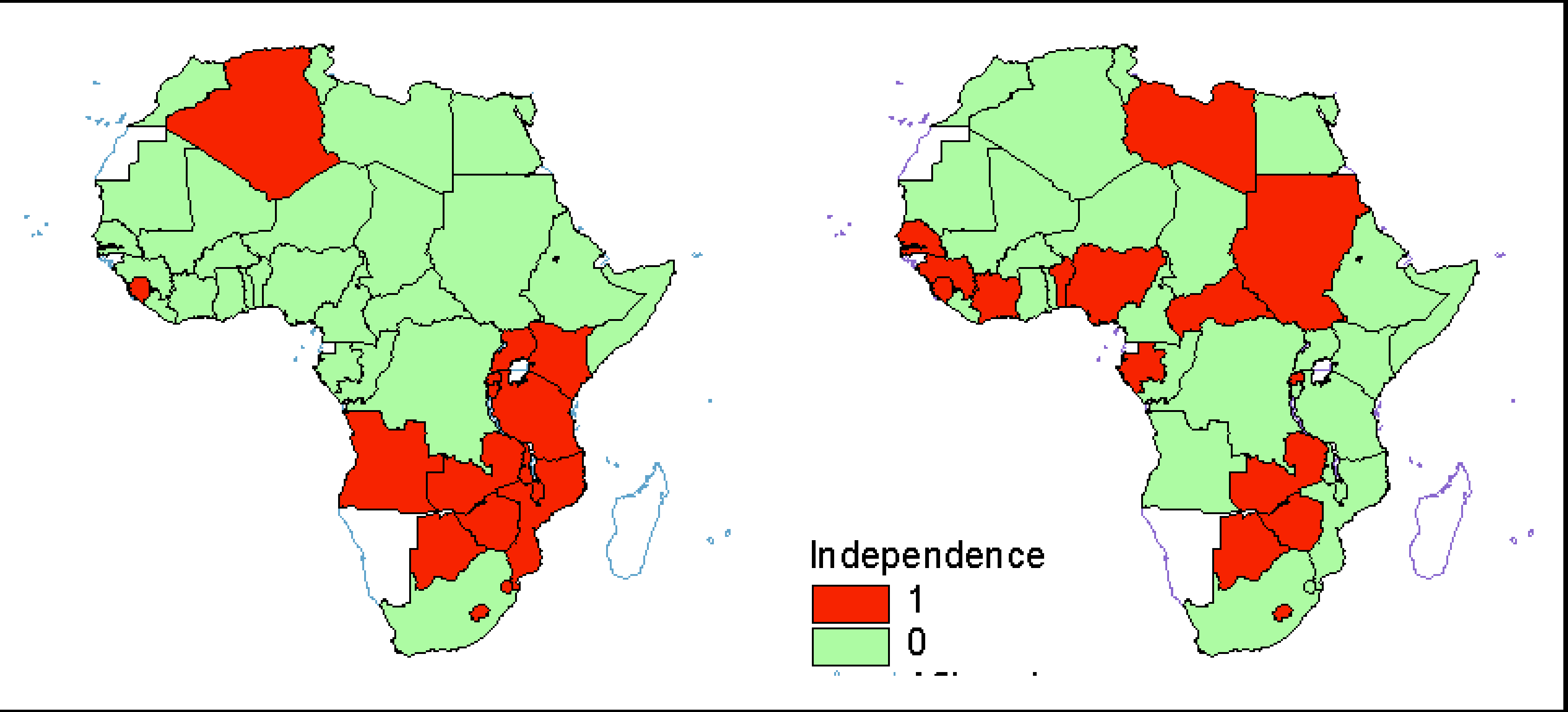
- Advantages
  - non-parametric
    - no distributional assumptions
  - easy to compute
  - easy to interpret
- Disadvantages
  - sample specific
    - no generalization to population
  - precision of pseudo significance arbitrary
    - 1/(99+1) yields 0.01,1/(999+1) yields 0.001
  - sensitive to random number generator

Join Count Statistic

Join Count Statistics

- Binary Data
- Match Value-Location
  - count joins in value that match joins in space = neighbors
  - x_i = 1 or 0, B(lack) or W(hite)
- Types of Joins
  - BB 1—1 $x_i.x_j$
  - WW 0—0 $(1-x_i)(1-x_j)$
  - BW 1—0 $(x_i - x_j)^2$
    - three types exhaust sample

Observed (left) and randomized (right) distribution for African Independence 60



BB = 18

BB = 9

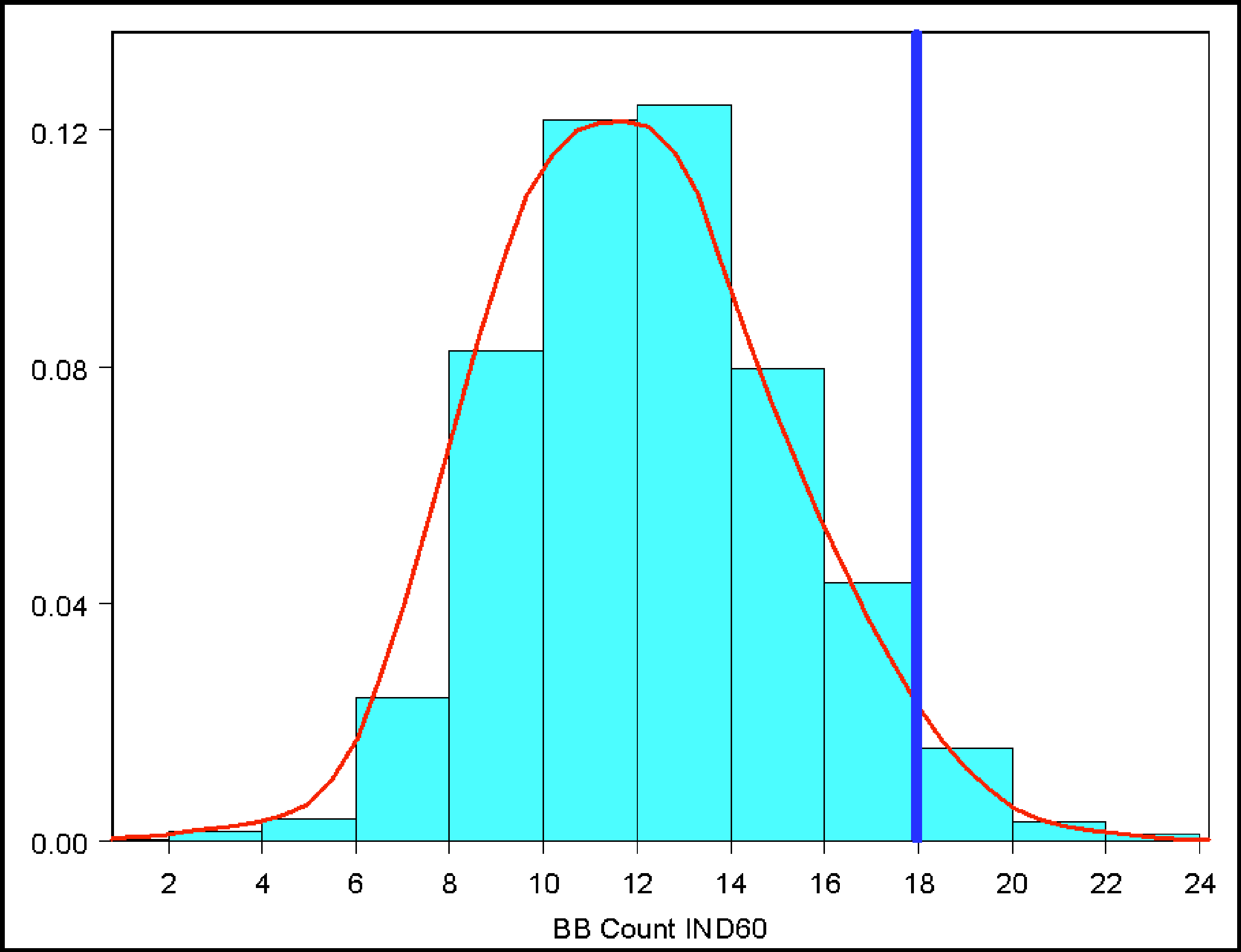
Classic Join Count Statistics

- Statistics
  - BB: $(1/2)\ \Sigma_i\ \Sigma_j\ w_{ij}.x_i.x_j$
  - WW: $(1/2)\ \Sigma_i\ \Sigma_j\ w_{ij}.(1-x_i).(1-x_j)$
  - BW: $(1/2)\ \Sigma_i\ \Sigma_j\ w_{ij}.(x_i - x_j)^2$
  - $BB + WW + BW = (1/2)\ S_0$
    $= (1/2)\ \Sigma_i\ \Sigma_j\ w_{ij}$
- Weights
  - $w_{ij} = 1$ or 0 (binary, symmetric)

Inference for Join Counts

- Use Sampling Theory
  - binomial with or without replacement
  - free and non-free sampling
- Randomization/Permutation
  - compute reference distribution
  - assess "extreme" value
  - pseudo-significance

Reference distribution for BB Count
(999 permutations)



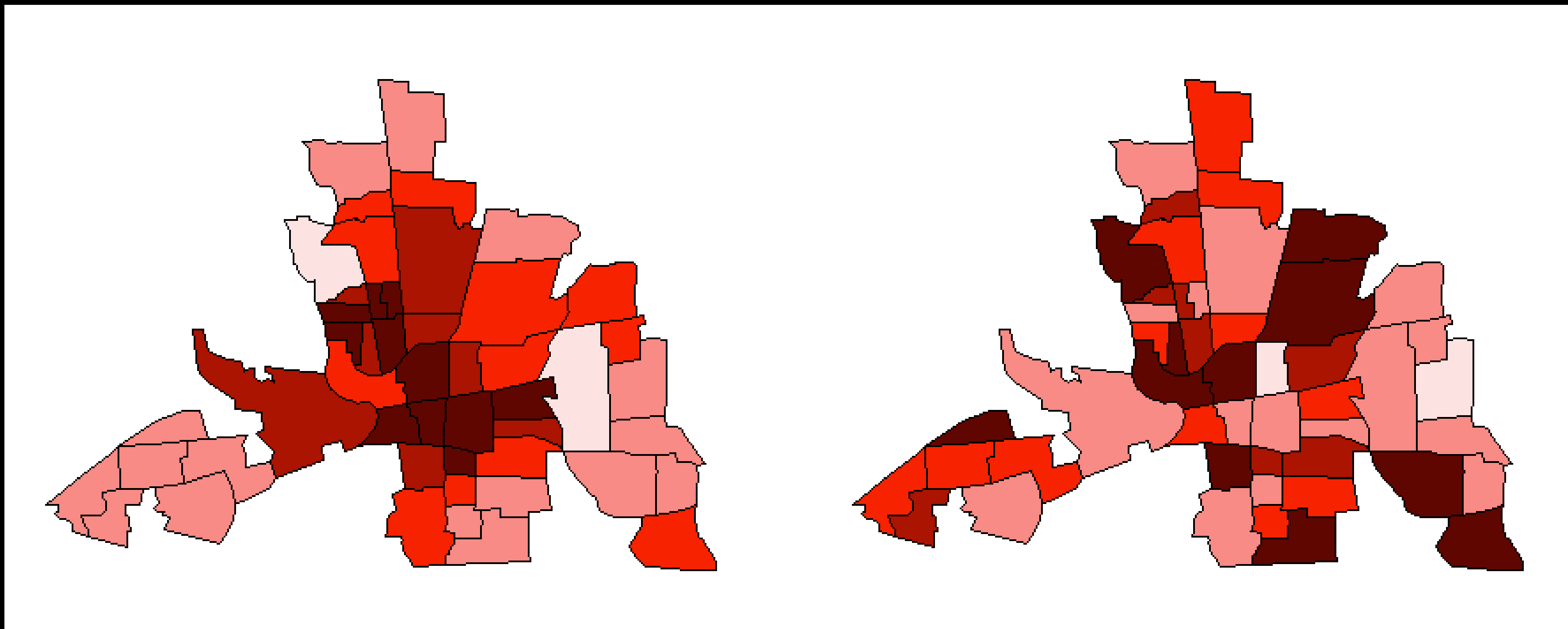
P[BB ≥ 18] = 0.04

Moran’s I

Moran's I

- Moran's *I* Spatial Autocorrelation Statistic
  - cross-product statistic
    $I = (N/S_0)\ \Sigma_i\ \Sigma_j\ w_{ij} .\ z_i . z_j\ /\ \Sigma_i\ z_i^2$
    with $z_i = x_i - \mu$ and $S_0 = \Sigma_i\ \Sigma_j\ w_{ij}$
- Inference
  - normal distribution
  - randomization
  - permutation

Observed (left) and randomized (right) distribution for Columbus Crime

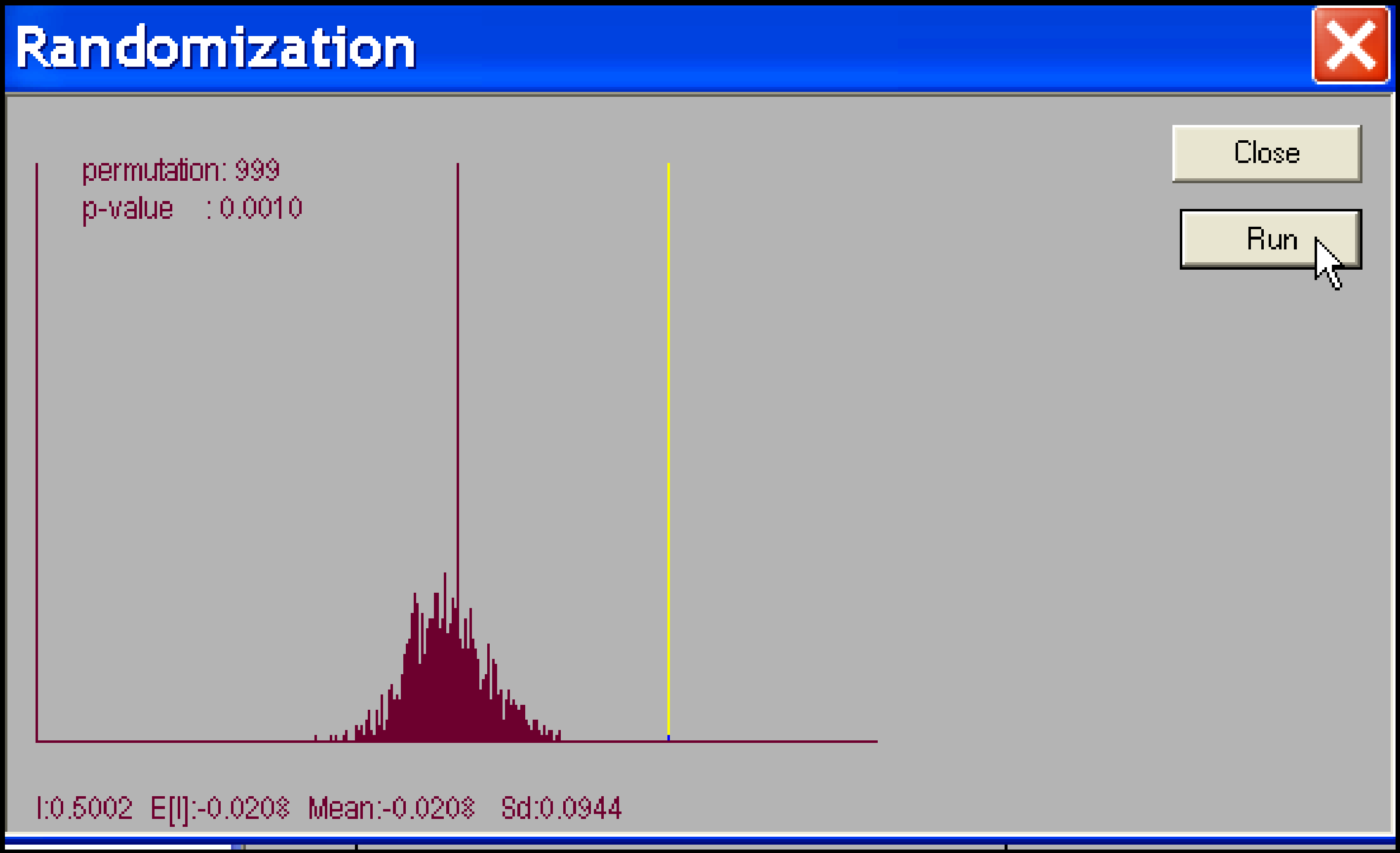


Moran's I = 0.486

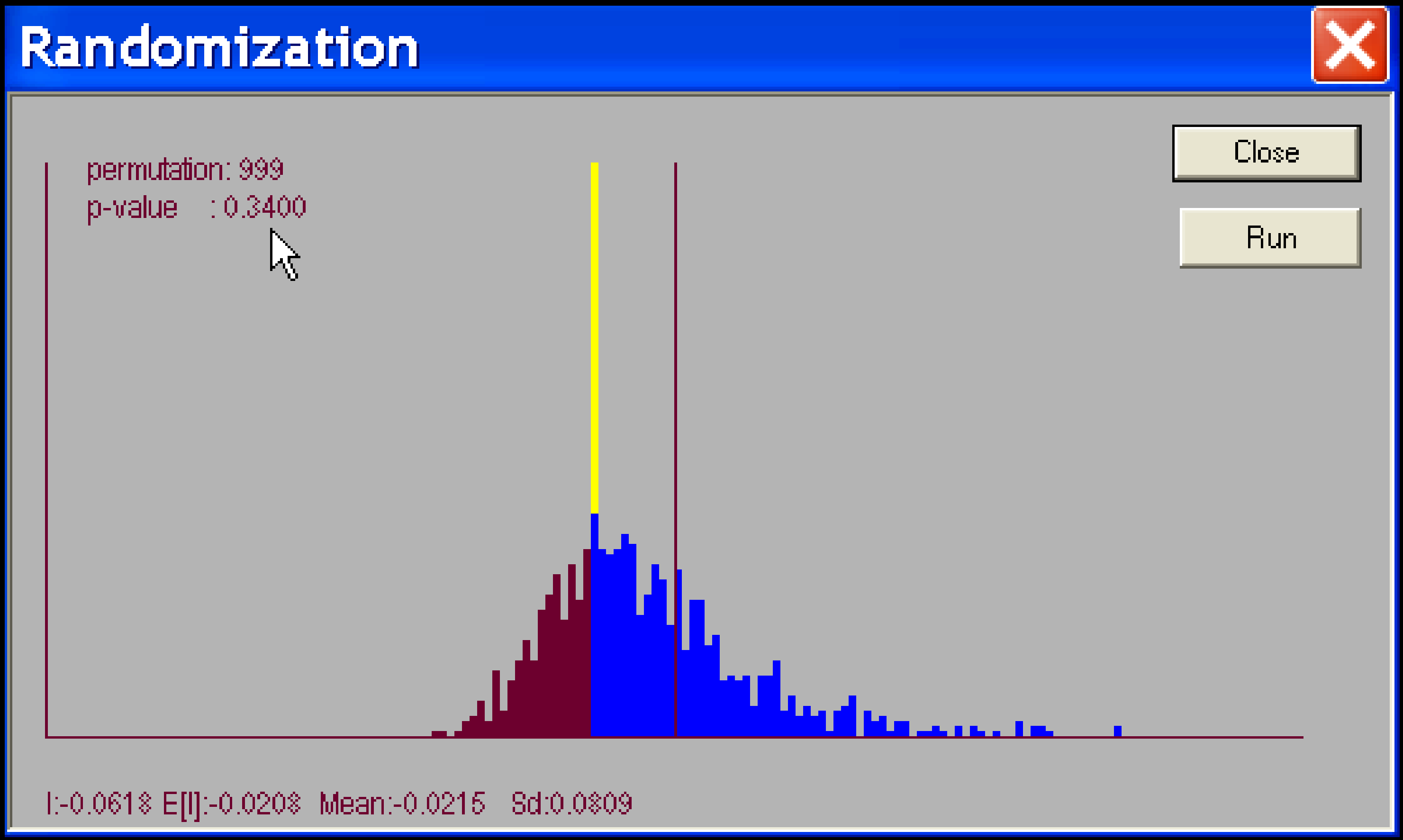
	0.178 - 0.224
	0.224 - 26.645
	26.645 - 39.175
	39.175 - 50.732
	50.732 - 68.892

Moran's I = -0.003

Reference Distribution (CRIME)



Reference Distribution (OPEN)



Interpretation of Moran’s I

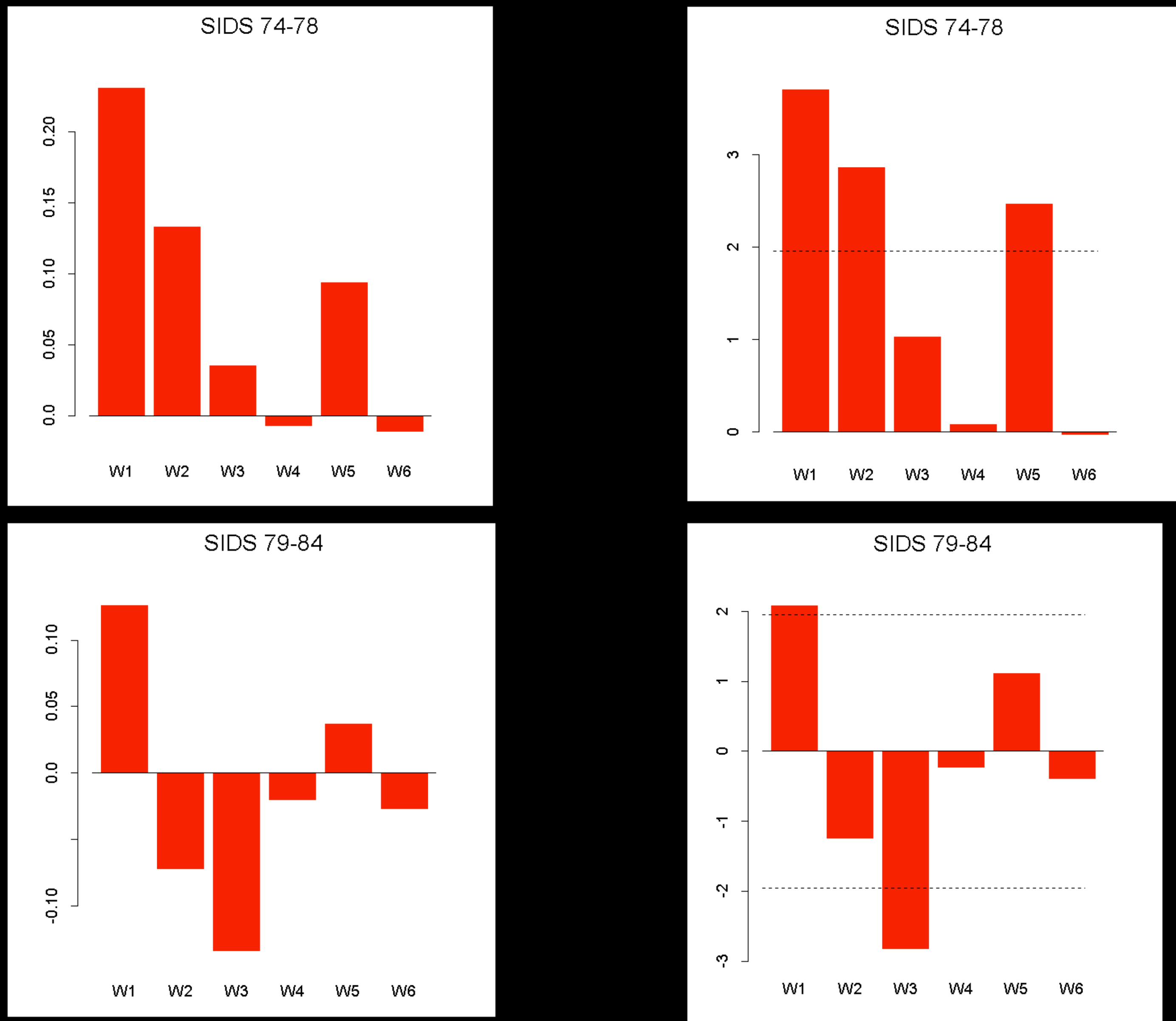
- Positive Spatial Autocorrelation
  - $I > -1/(n-1)$, or $z > 0$
  - spatial clustering of high and/or low values
    - no distinction between high or low
- Negative Spatial Autocorrelation
  - $I < -1/(n-1)$, or $z < 0$
  - checkerboard pattern, “competition”

Moran's I for Rates

- Variance Instability
  - violates assumption of stationarity
  - spurious spatial correlation
- EB Adjustment (Assuncao-Reis)
  - standardize each rate
    - $z_i = (r_i - R/P) / SE$
  - use standardized rate

Spatial (lag) Correlogram

- Visualization
  - spatial autocorrelation statistics for increasing orders of contiguity
  - use z-values
- Interpretation
  - identification of spatial process
  - range of association
    - possible indication of misspecified spatial weights and/or nonstationarity



Moran's I by W

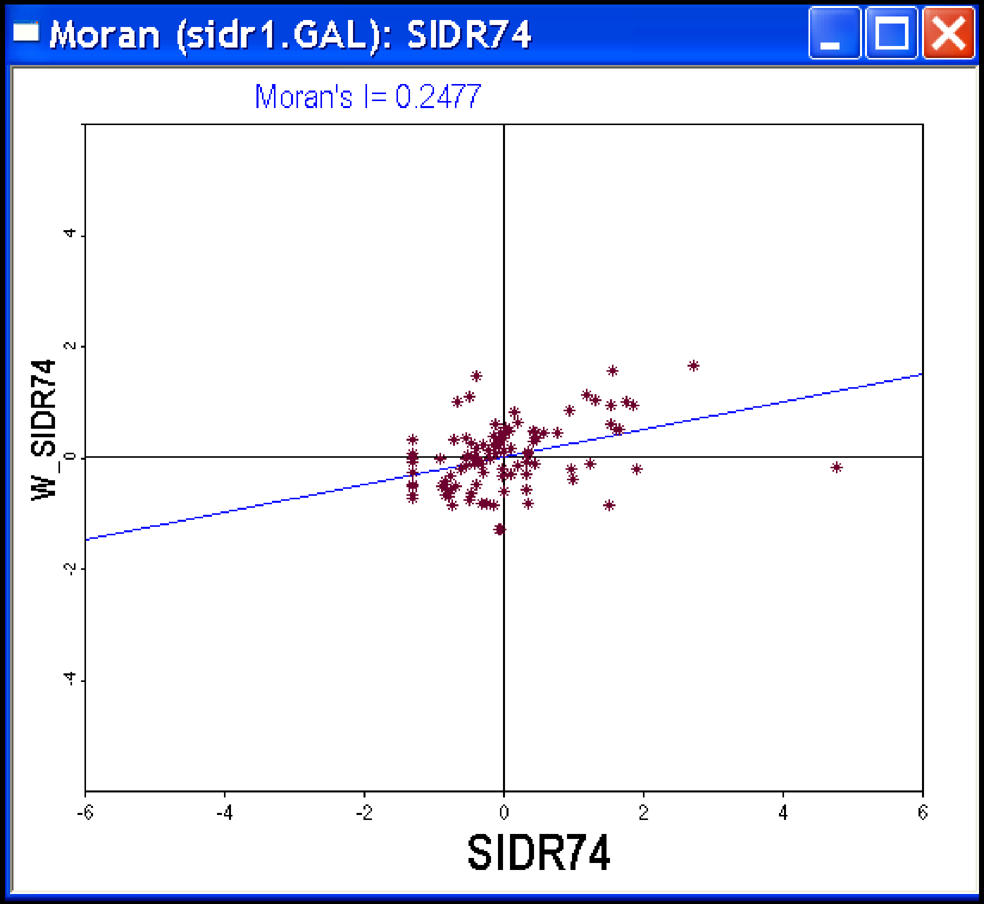
z-value by W

Moran Scatterplot

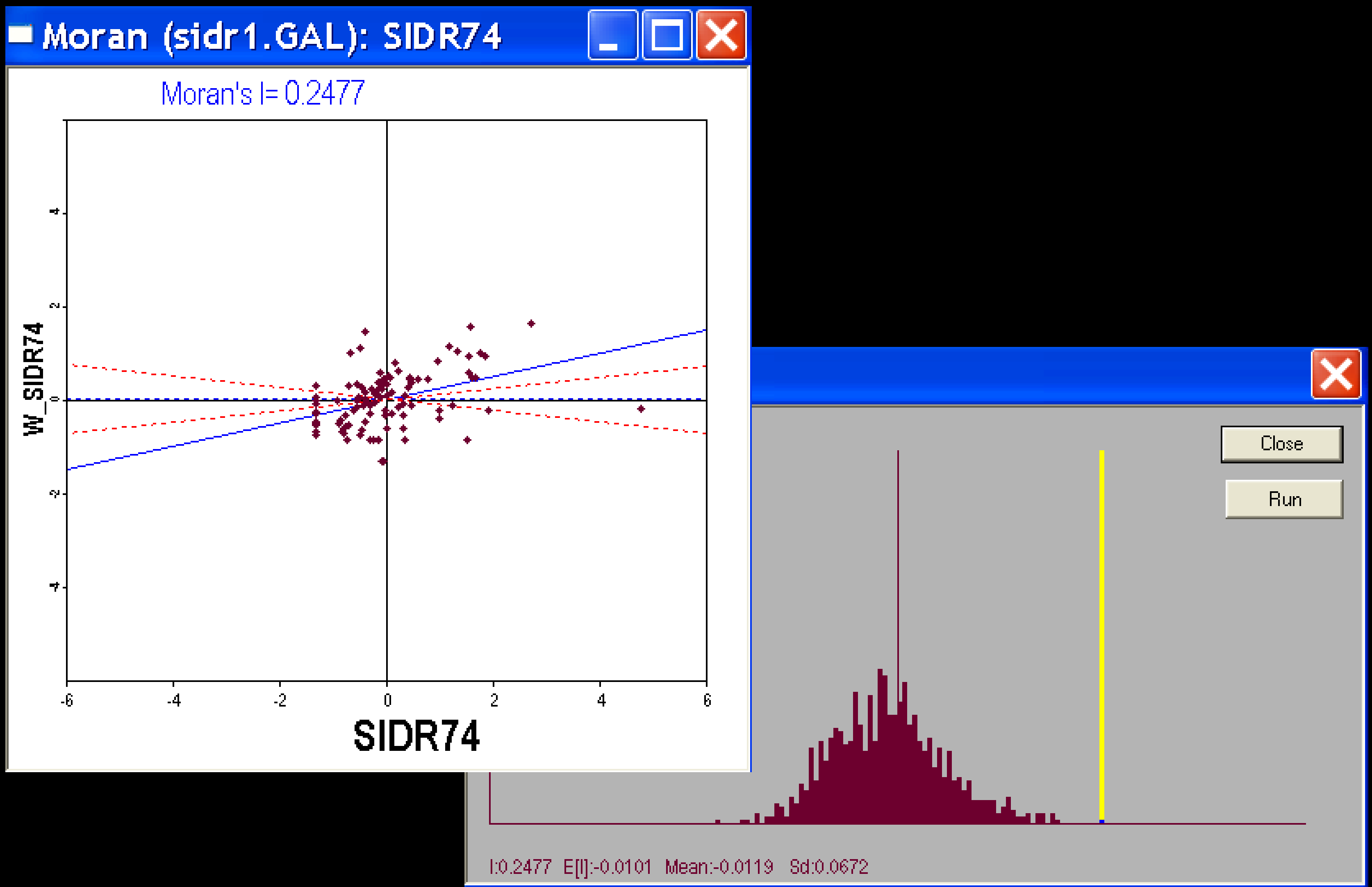
Moran Scatterplot

- Linear Spatial Autocorrelation
  - linear association between value at i and weighted average of neighbors: $\Sigma_j w_{ij} y_j$ vs. y_i , or Wy vs y
  - four quadrants
    - high-high, low-low = spatial clusters
    - high-low, low-high = spatial outliers
- Moran's I
  - slope of linear scatterplot smoother
  - $I = z'Wz / z'z$

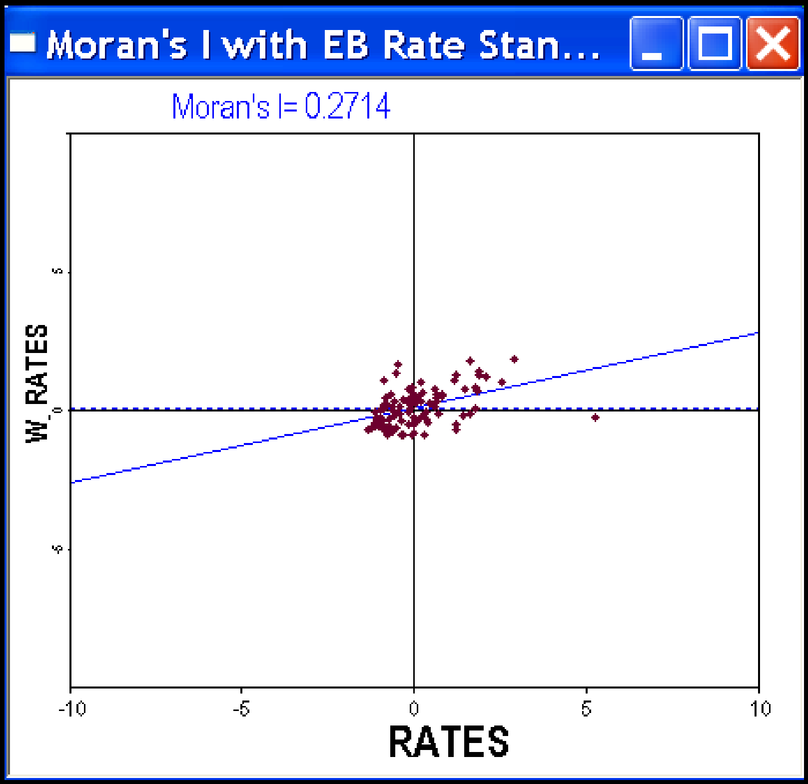
Moran Scatterplot

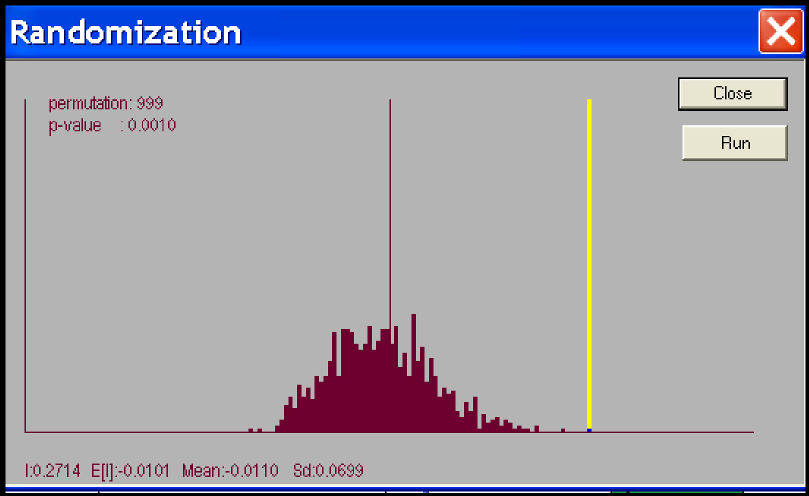


Significance Envelope



EB Adjusted Moran Scatterplot

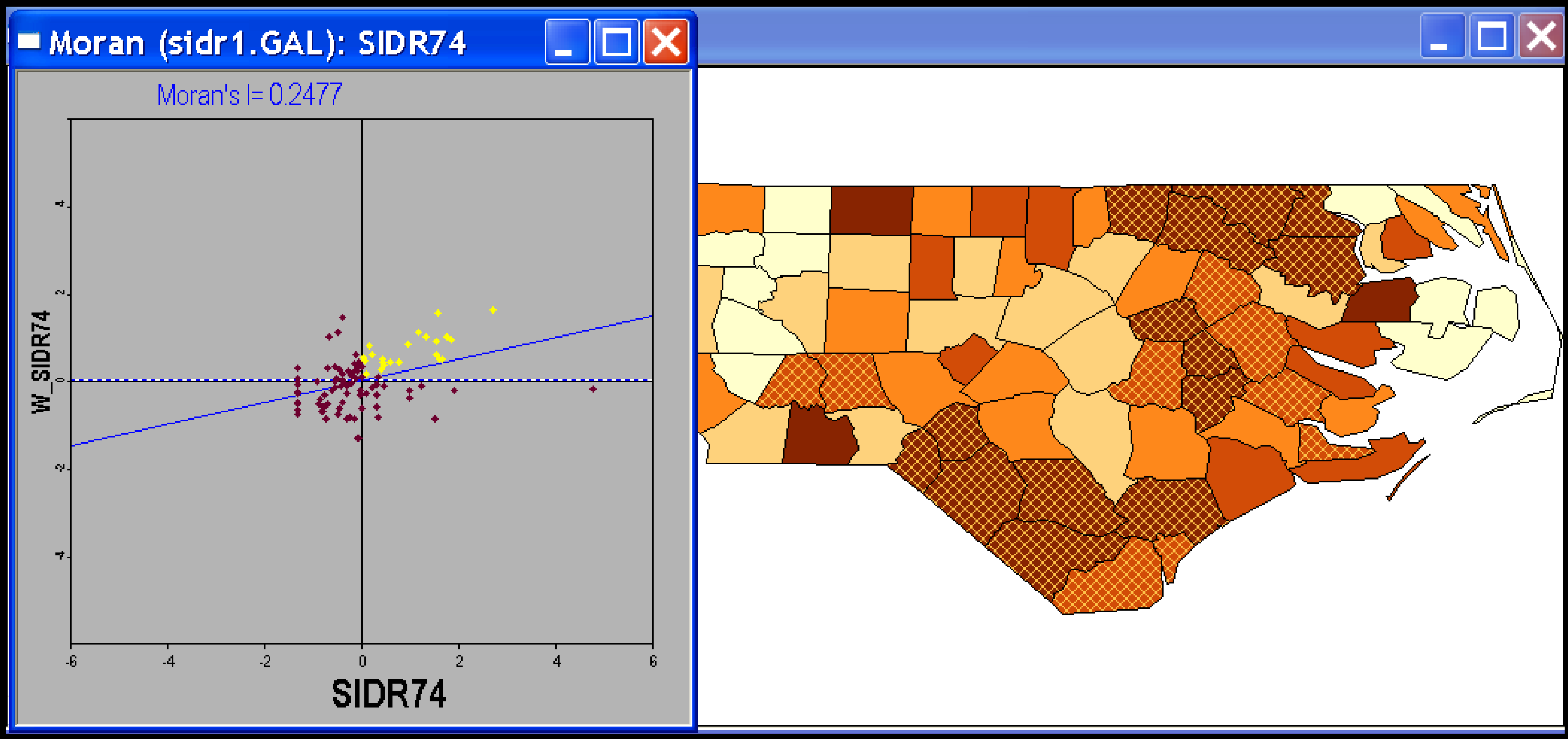




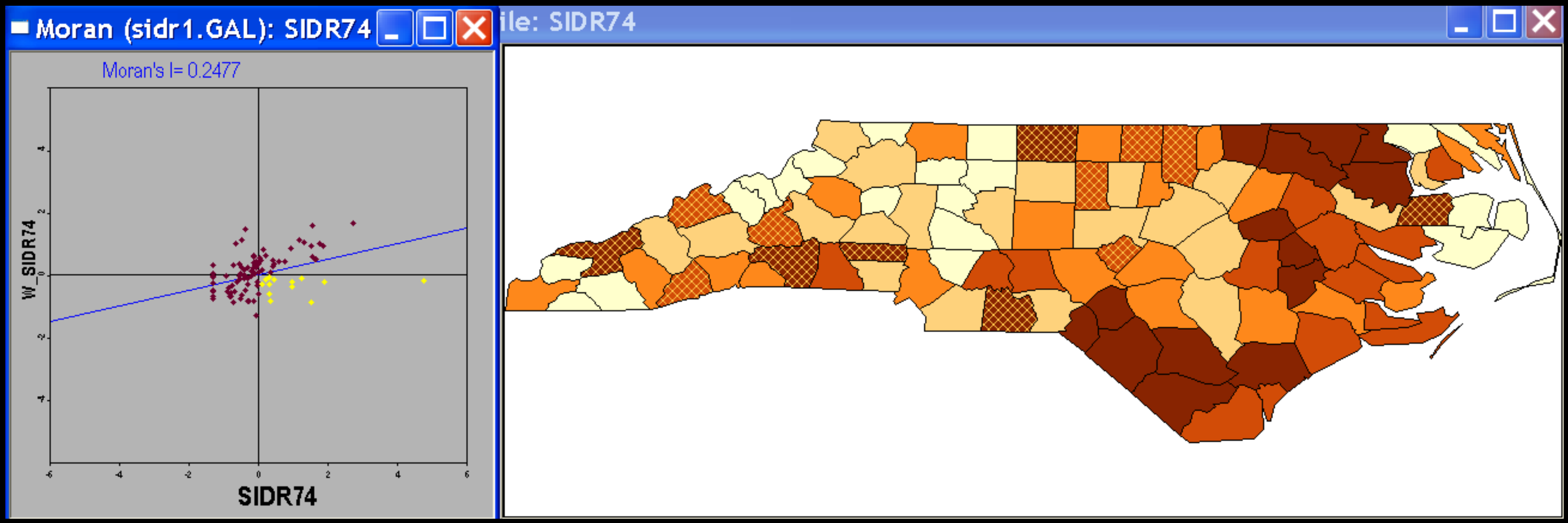
Use of Moran Scatterplot

- Classification of Spatial Autocorrelation
- Local Nonstationarity
  - outliers
  - high leverage points
  - sensitivity to boundary values
- Regimes
  - different slopes in subsets of the data

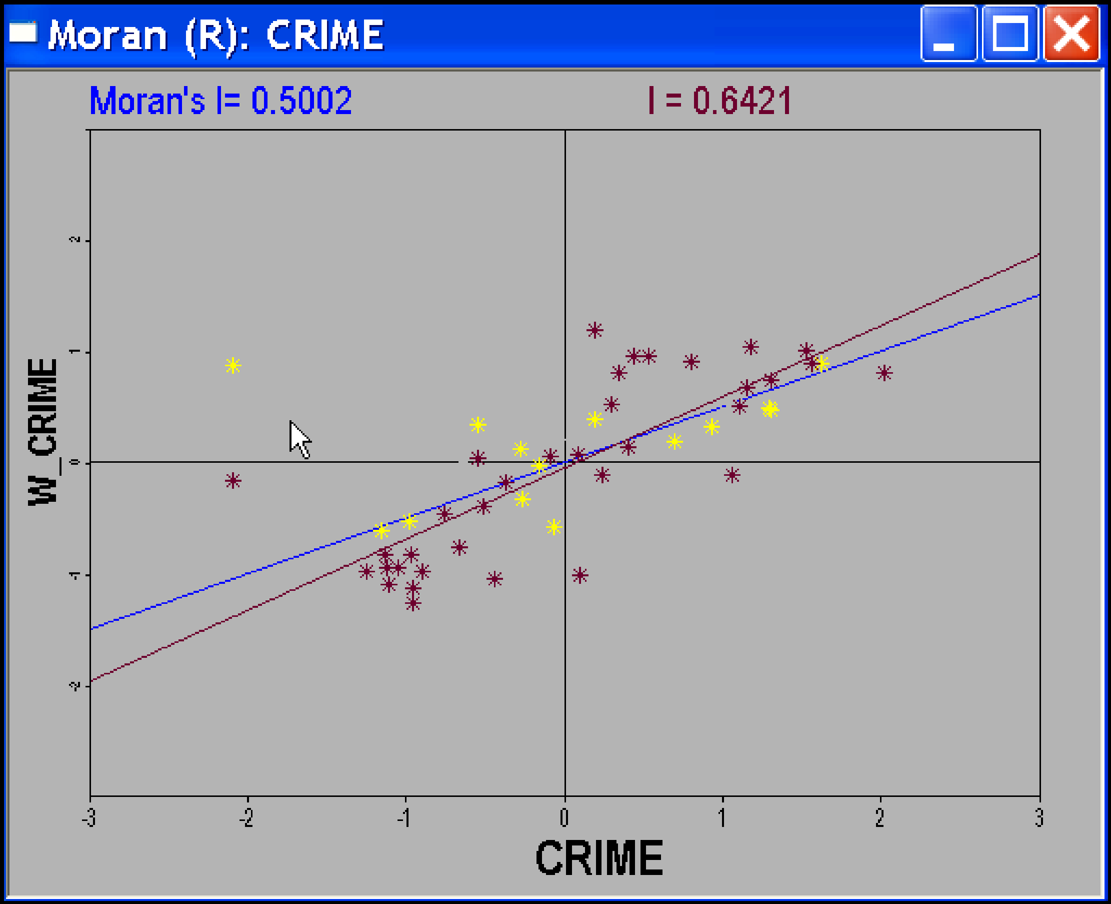
Spatial Clusters



Spatial Outliers



Moran Scatterplot - Regimes

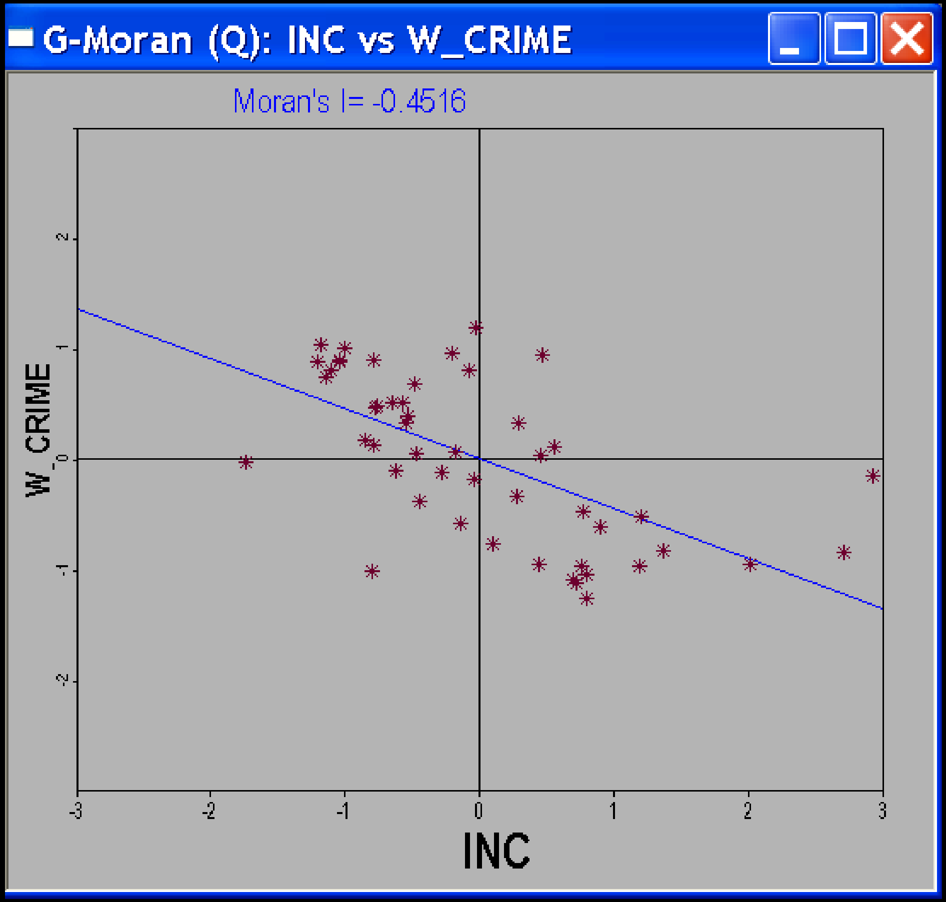


Bivariate/Space-Time

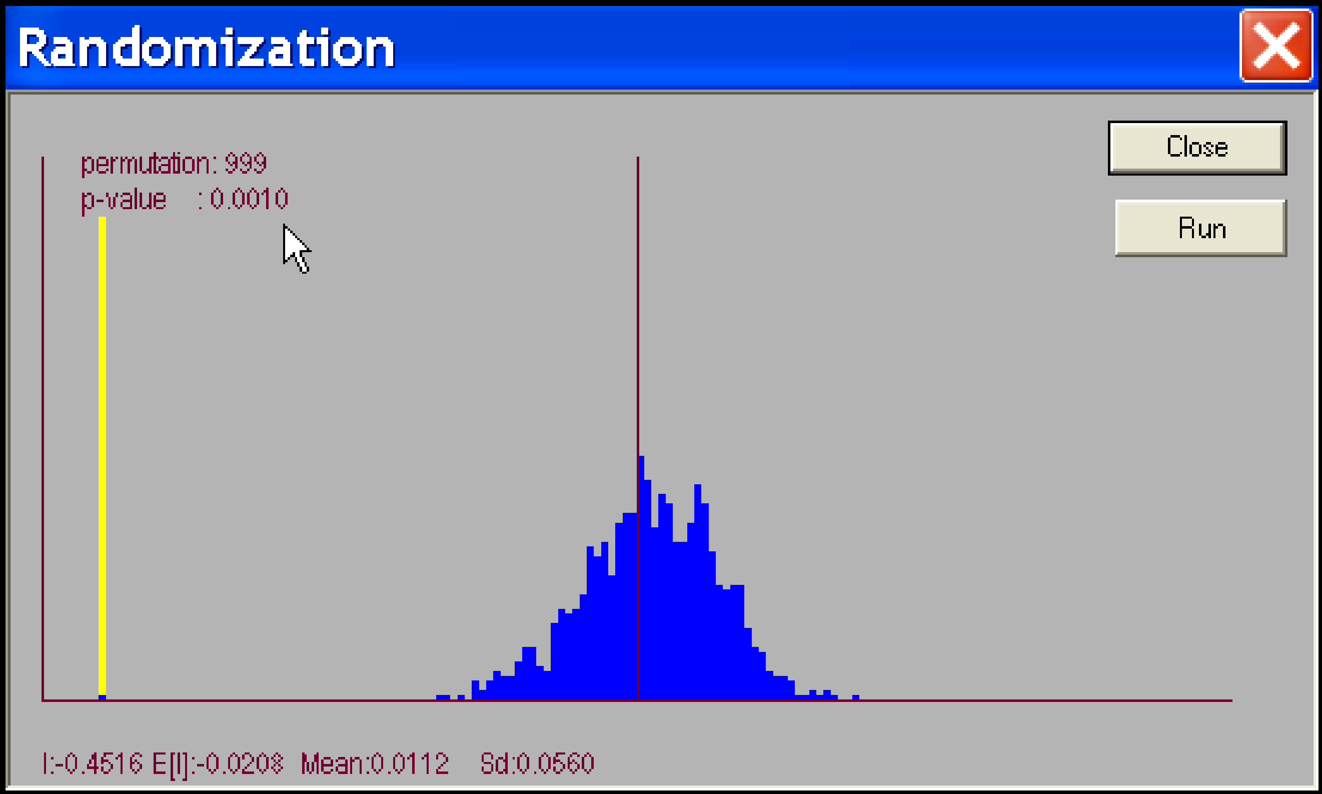
Moran Scatterplot Extensions

- Generalized Moran Scatterplot
  - Regression slope of Wz_2 on z_1
    - both variables standardized
    - = visualization of multivariate Moran statistic (Wartenberg)
  - Significance testing
    - permutation
    - permutation envelope (2.5% and 97.5% from permutation reference distribution)
- Four Types of Association
  - High-high, Low-low; High-low, Low-high

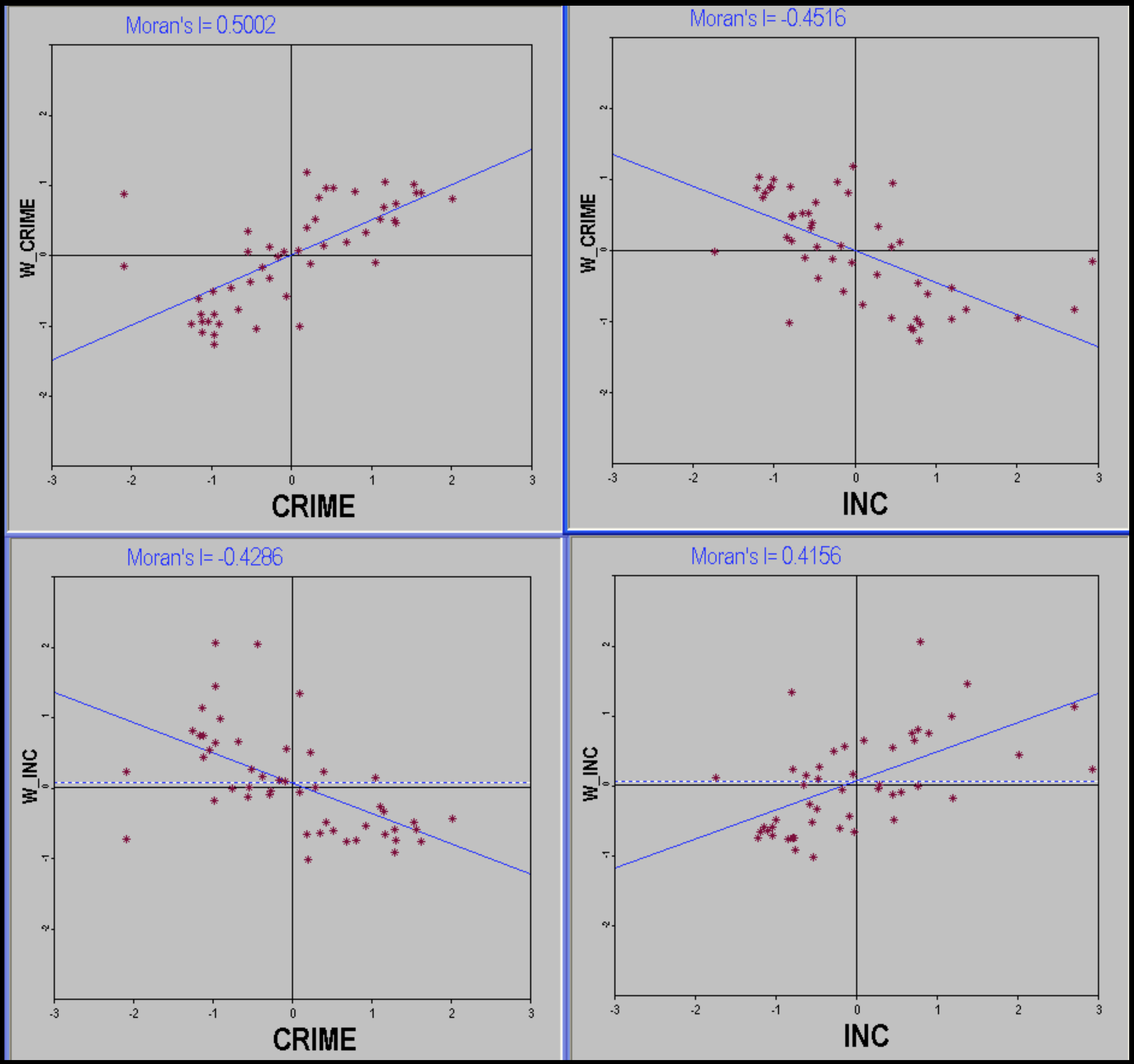
Bivariate Moran



Reference Distribution Bivariate Moran



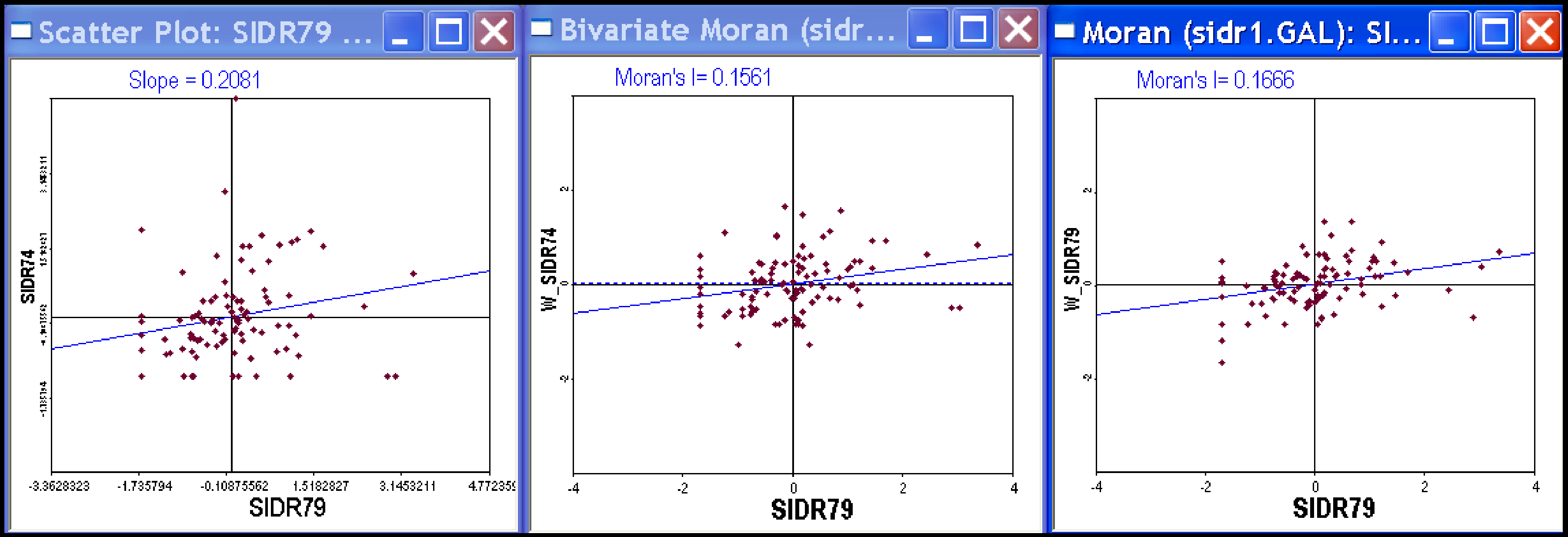
Moran Scatterplot Matrix



Space-Time Association

- Serial Autocorrelation
  - z at i,t with z at i,t+1
    - standard scatterplot
- Spatial Autocorrelation
  - z at i,t with z at j,t
  - z at i, t+1 with z at j,t+1
    - Moran scatterplot
- Space-Time Correlation
  - z at i,t with z at j,t+1
    - bivariate Moran scatterplot

Space-Time Scatterplots



Geary’s c Statistic

Geary's c Spatial Autocorrelation Statistic

- Geary's c
  - squared differences
    $c = (N-1).\ \Sigma_i\ \Sigma_j\ w_{ij}.(x_i-x_j)^2$
    $/\ 2(S_0).\ \Sigma_i\ z_i^2$
    with $z_i = x_i - \mu$ and $S_0 = \Sigma_i\ \Sigma_j\ w_{ij}$
- Inference
  - normal distribution
  - randomization
  - permutation

Moments of Geary's c

- Normal approximation
  - $E[c] = 1$
    - does not depend on W or y, nor on sample size
  - $Var[c] = [(2S_1 + S_2)(n-1) - 4S_0^2] / [2(n+1)S_0^2]$
    - normal approximation does not depend on r.v., randomization approximation does
- Inference
  - $z = \{ c - E[c] \} / SD[c]$
  - normal approximation

Interpretation of Geary's c

- Positive Spatial Autocorrelation
  - $c < 1$, or $z < 0$
  - spatial clustering of high and/or low values
    - no distinction between high or low
    - opposite sign from Moran's I
- Negative Spatial Autocorrelation
  - $c > 1$, or $z > 0$
  - checkerboard pattern, "competition"

[redacted]