• Positive Spatial Autocorrelation Does Not Imply Diffusion

• diffusion tends to yield positive spatial autocorrelation, but the reverse is not necessary

• spatial correlation may be due to structural factors, without contagion or diffusion

• What is the Cause Behind Clustering

• True Contagion

• the result of a contagious process, social interaction,adynamicprocess

Apparent Contagion

• • spatial heterogeneity, stratification

Cannot Be Distinguished In a Pure Cross Section

Clustering

• Global Characteristic of Spatial

Pattern - NOT Local

• are like values more grouped in space

• than random

property of overall pattern = all the

• observations

test by means of a global spatial

• autocorrelation statistic

Clusters

• Local Characteristic of Spatial Pattern • where are like values more grouped in

• space than random

property of local pattern = location-

• specific

test by means of a local spatial

• autocorrelation statistic

local clusters may be compatible with global spatial randomness

Spatial Autocorrelation Statistic

• Formal Test of Match Between Value • Similarity and Locational Similarity • Statistic Summarizes Both Aspects

Significance

• how likely is it (p-value) that the computed statistic would take this (extreme) value in a spatially random pattern

Spatial autocorrelation

est for the presence of spatial autocorrelation

Global

Moran’sI

Geary’s C

Local (LISA – Local Indicators of Spatial Autocorrelation)

LocalMoran’sIandGetisGi\*

Other tests that are more simple:

The Chi‐square Test for Spatial Independence

The Join Count Statistic

Scaling Factors

Moran’s I (2)

• denominator: N = number of • observations

• numerator: S0 = Σi Σj wij

S0

• number of nonzero elements in W

• number of neighbor pairs (x2)

• Normal

•• assume uncorrelated normal distribution

Z-Value

• compute E[I] and Var[I] • z = (I - E[I]) / SD[I]

• Permutation Approach

• reshuffle observations

• construct reference distribution from random permutations

• pseudo significance

• p = ( M + 1) / (N + 1)

• For Significant Statistics Only

• Use z-value

•• I depends on W

Positive S.A. zI > 0 for p < 0.05 ...

• no distinction between clustering of high • or low values

Negative S.A. zI < 0 for p < 0.05 ...

• Variance Instability of Rates • non-constant variance violates • assumption of stationarity

may lead to spurious indication of spatial • autocorrelation

Empirical Bayes Adjustment

(Assuncao-Reis)

• standardize each rate • use standardized rate

•

•• is slope in regressionWz = a + I.z Moran Scatter Plot

• •

Moran’s I as a Regression Slope

• in matrix notation: I = z’Wz / z’z

linear association between Wz on the y- axis and z on the x-axis

each point is pair (zi,Wzi), slope is I

• •

Negative Spatial Autocorrelation

• • high-low and low-high: spatial outliers

Only Suggestive

Four Categories of SA

Positive Spatial Autocorrelation

• • high-high and low-low: spatial clusters

•

no suggestion of significance

•

relative to mean

Classic/best measure of spatial autocorrelation

Depends upon definition of neighboring unit via the

spatial weights matrix

Typically ranges from -1 to 1

Like regression, it has a few assumptions

Regional x/y values all come from normal distributions w/same

mean and variance for each region

Randomly rearrange the data on map and compute I many times, would have a normal distribution

Why? Because we use the normal distribution to calculate the p-value

What if my data violate the assumptions?

If you doubt that the assumptions of Moran’s I are true (normality and randomization), we can use a Monte Carlo simulation

Simulate Moran’s I n times under the assumption of no spatial

pattern

Assigning all regions the mean value

CalculateMoran’sI

Compare actual value of Moran’s I to randomly simulated distribution to obtain p-value

Global vs. Local Analysis

Global

One statistic to summarize pattern in whole study area

Clustering

Homogeneity

LISA: local indicators of spatial association

Location‐specific statistics

Clusters/hot‐spots

Heterogeneity

Anselin (1995)

•

Local Spatial Statistic

LISA Definition

•

Local-Global Relation

indicate significant spatial autocorrelation for each location

•

sum of LISA proportional to a

•

corresponding global indicator of spatial

autocorrelation

Geary’s C

Geary's C typically ranges from 0 to 3 Cannot be negative

An uncorrelated process has an expected C = 1

Values less than 1 indicate positive spatial autocorrelation

Values greater than 1 indicate negative autocorrelation

Relationship of Moran’s I and Geary’s C

C approaches 0 and I approaches 1 when similar values are clustered

C approaches 3 and I approaches -1 when dissimilar values tend to cluster

High values of C measures correspond to low values of I

So the two measures are inversely related

Local Moran’s I

Used to determine if local autocorrelation exists around each region I

Returns an I value for each region

Often used after global Moran’s I to see if:

The study area is homogeneous (local statistics similar across regions)

There are local outliers that contribute to a significant global statistic

Spatial weights matrices

Neighborhoods can be defined in a number of ways Contiguity (common boundary)

What is a “shared” boundary?

Distance (distance band, K-nearest neighbors)

How many “neighbors” to include, what distance do we use? General weights (social distance, distance decay)

Most common is using binary connectivity based on contiguity

wij = 1 if regions i and j are contiguous, wij = 0 otherwise

May also be defined as a function of the distance between i and j

Distance of the line connecting the centroids of two areas

wij = dij-

wij = exp[-dij]

A binary weights matrix looks like: 0 1 0 0 0011 1100 0111

Observation 1 has neighbor 2

Observation 2 has neighbors 3 and 4

Observation 3 has neighbors 1 and 2

Observation 4 has neighbor 2, 3 and 4

A row-standardized matrix it looks like: 0 1 0 0

0 0

0 .33 .33 .33