Spatial Analysis and Modeling
(GIST 4302/5302)

Guofeng Cao
Department of Geosciences
Texas Tech University
Outline of This Week

• Last week, we learned:
  – spatial point pattern analysis (PPA)
  – focus on location distribution of ‘events’
  – Measure the cluster (spatial autocorrelation) in point pattern

• This week, we will learn:
  – How to measure and detect clusters/spatial autocorrelation in areal data (regional data)
Spatial Autocorrelation

• Spatial autocorrelation is everywhere
  – Spatial point pattern
    • K, F, G functions
    • Kernel functions
  – Areal/lattice (this topic)
  – Geostatistical data (next topic)
Spatial Autocorrelation of Areal Data
Spatial Autocorrelation

- Tobler’s first law of geography
- **Spatial auto/cross correlation**

If like values tend to cluster together, then the field exhibits high **positive spatial autocorrelation**.

If there is no apparent relationship between attribute value and location then there is **zero spatial autocorrelation**.

If like values tend to be located away from each other, then there is **negative spatial autocorrelation**.
Positive spatial autocorrelation

- high values surrounded by nearby high values
- intermediate values surrounded by nearby intermediate values
- low values surrounded by nearby low values

Source: Ron Briggs of UT-Dallas
**Negative spatial autocorrelation**

- high values surrounded by nearby low values
- intermediate values surrounded by nearby intermediate values
- low values surrounded by nearby high values

Source: Ron Briggs of UT Dallas
Measuring Spatial Autocorrelation: the problem of measuring “nearness”

To measure spatial autocorrelation, we must know the “nearness” of our observations as we did for point pattern case

- Which points or polygons are “near” or “next to” other points or polygons?
  - Which states are near Texas?
  - How to measure this?

Seems simple and obvious, but it is not!
Spatial Weight Matrix

• **Core** concept in statistical analysis of areal data

• Two steps involved:
  – define which relationships between observations are to be given a nonzero weight, i.e., define spatial neighbors
  – assign weights to the neighbors
Spatial Neighbors

• **Contiguity-based neighbors**
  – Zone \( i \) and \( j \) are neighbors if zone \( i \) is contiguity or adjacent to zone \( j \)
  – But what constitutes contiguity?

• **Distance-based neighbors**
  – Zone \( i \) and \( j \) are neighbors if the distance between them are less than the threshold distance
  – But what distance do we use?
Contiguity-based Spatial Neighbors

• Sharing a border or boundary
  – Rook: sharing a border
  – Queen: sharing a border or a point

rook

queen

Hexagons

Irregular

Which use?
“Close” but no common border

Length of border
- Is Arizona “as close to” California as to Utah?
- Base “closeness” on proportion of shared border, not just one (1) or zero (0)
- \( w_{ij} = \frac{\text{border length}_{ij}}{\text{border length}_j} \)
Higher-Order Contiguity

1\textsuperscript{st} order

Nearest neighbor

rook

2\textsuperscript{nd} order

Next nearest neighbor

hexagon

queen
Distance-based Neighbors

• How to measure distance between polygons?

• Distance metrics
  – 2D Cartesian distance (projected data)
  – 3D spherical distance/great-circle distance (lat/long data)
    • Haversine formula

\[
\text{Haversine formula: } a = \sin^2(\Delta \varphi/2) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2(\Delta \lambda/2) \\
c = 2 \cdot \arctan2(\sqrt{a}, \sqrt{1-a}) \\
d = R \cdot c
\]

where \( \varphi \) is latitude, \( \lambda \) is longitude, \( R \) is earth’s radius (mean radius = 6,371km)
Distance-based Neighbors

- k-nearest neighbors

Fig. 9.5. (a) $k = 1$ neighbours; (b) $k = 2$ neighbours; (c) $k = 4$ neighbours

Source: Bivand and Pebesma and Gomez-Rubio
Distance-based Neighbors

- thresh-hold distance (buffer)

Fig. 9.6. (a) Neighbours within 1,158 m; (b) neighbours within 1,545 m; (c) neighbours within 2,317 m

Source: Bivand and Pebesma and Gomez-Rubio
Neighbor/Connectivity Histogram

Source: Bivand and Pebesma and Gomez-Rubio
Spatial Weight Matrix

• Spatial weights can be seen as a list of weights indexed by a list of neighbors
• If zone j is not a neighbor of zone i, weights Wij will set to zero
  – The weight matrix can be illustrated as an image
  – Sparse matrix
A Simple Example for Rook case

- Matrix contains a:
  - 1 if share a border
  - 0 if do not share a border

4 areal units

Common border

4x4 matrix

\[
\begin{array}{cccc}
A & B & C & D \\
A & 0 & 1 & 1 & 0 \\
B & 1 & 0 & 0 & 1 \\
C & 1 & 0 & 0 & 1 \\
D & 0 & 1 & 1 & 0 \\
\end{array}
\]
<table>
<thead>
<tr>
<th>Name</th>
<th>Fips</th>
<th>Ncount</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1</td>
<td>4</td>
<td>28</td>
<td>13</td>
<td>12</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>4</td>
<td>5</td>
<td>35</td>
<td>8</td>
<td>49</td>
<td>6</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>5</td>
<td>6</td>
<td>22</td>
<td>28</td>
<td>48</td>
<td>47</td>
<td>40</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>32</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>8</td>
<td>7</td>
<td>35</td>
<td>4</td>
<td>20</td>
<td>40</td>
<td>31</td>
<td>49</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>9</td>
<td>3</td>
<td>44</td>
<td>36</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>10</td>
<td>3</td>
<td>24</td>
<td>42</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>11</td>
<td>2</td>
<td>51</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>12</td>
<td>2</td>
<td>13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>13</td>
<td>5</td>
<td>12</td>
<td>45</td>
<td>37</td>
<td>1</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td>16</td>
<td>5</td>
<td>32</td>
<td>41</td>
<td>56</td>
<td>49</td>
<td>30</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>17</td>
<td>5</td>
<td>29</td>
<td>21</td>
<td>18</td>
<td>55</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>18</td>
<td>4</td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>19</td>
<td>6</td>
<td>29</td>
<td>31</td>
<td>17</td>
<td>55</td>
<td>27</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>20</td>
<td>4</td>
<td>40</td>
<td>29</td>
<td>31</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>21</td>
<td>7</td>
<td>47</td>
<td>29</td>
<td>18</td>
<td>39</td>
<td>54</td>
<td>51</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>22</td>
<td>3</td>
<td>28</td>
<td>48</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>23</td>
<td>1</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>24</td>
<td>5</td>
<td>51</td>
<td>10</td>
<td>54</td>
<td>42</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>25</td>
<td>5</td>
<td>44</td>
<td>9</td>
<td>36</td>
<td>50</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>26</td>
<td>3</td>
<td>18</td>
<td>39</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>27</td>
<td>4</td>
<td>19</td>
<td>55</td>
<td>46</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td>28</td>
<td>4</td>
<td>22</td>
<td>5</td>
<td>1</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td>29</td>
<td>8</td>
<td>5</td>
<td>40</td>
<td>17</td>
<td>21</td>
<td>47</td>
<td>20</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>Montana</td>
<td>30</td>
<td>4</td>
<td>16</td>
<td>56</td>
<td>38</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td>31</td>
<td>6</td>
<td>29</td>
<td>20</td>
<td>8</td>
<td>19</td>
<td>56</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>32</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>49</td>
<td>16</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>33</td>
<td>3</td>
<td>25</td>
<td>23</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>34</td>
<td>3</td>
<td>10</td>
<td>36</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>35</td>
<td>5</td>
<td>48</td>
<td>40</td>
<td>8</td>
<td>4</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>36</td>
<td>5</td>
<td>34</td>
<td>9</td>
<td>42</td>
<td>50</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>37</td>
<td>4</td>
<td>45</td>
<td>13</td>
<td>47</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td>38</td>
<td>3</td>
<td>46</td>
<td>27</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>39</td>
<td>5</td>
<td>26</td>
<td>21</td>
<td>54</td>
<td>42</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>40</td>
<td>6</td>
<td>5</td>
<td>35</td>
<td>48</td>
<td>29</td>
<td>20</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>41</td>
<td>4</td>
<td>6</td>
<td>32</td>
<td>16</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>42</td>
<td>6</td>
<td>24</td>
<td>54</td>
<td>10</td>
<td>39</td>
<td>36</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>44</td>
<td>2</td>
<td>25</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>45</td>
<td>2</td>
<td>13</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>46</td>
<td>6</td>
<td>56</td>
<td>27</td>
<td>19</td>
<td>31</td>
<td>38</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>47</td>
<td>8</td>
<td>5</td>
<td>28</td>
<td>1</td>
<td>37</td>
<td>13</td>
<td>51</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>Texas</td>
<td>48</td>
<td>4</td>
<td>22</td>
<td>5</td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>49</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>35</td>
<td>56</td>
<td>32</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>50</td>
<td>3</td>
<td>36</td>
<td>25</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>51</td>
<td>6</td>
<td>47</td>
<td>37</td>
<td>24</td>
<td>54</td>
<td>11</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>53</td>
<td>2</td>
<td>41</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>54</td>
<td>5</td>
<td>51</td>
<td>21</td>
<td>24</td>
<td>39</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>55</td>
<td>4</td>
<td>26</td>
<td>17</td>
<td>19</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>56</td>
<td>6</td>
<td>49</td>
<td>16</td>
<td>31</td>
<td>8</td>
<td>46</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Style of Spatial Weight Matrix

• Row
  – a weight of unity for each neighbor relationship

• Row standardization
  – Symmetry not guaranteed
  – can be interpreted as allowing the calculation of average values across neighbors

• General spatial weights based on distances
# Row vs. Row standardization

## Total number of neighbors
--some have more than others

## Row standardized
--usually use this

### Divide each number by the row sum

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Row Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

### Row Sum

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Row Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0.3</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>1</td>
</tr>
</tbody>
</table>
General Spatial Weights Based on Distance

- Decay functions of distance
  - Most common choice is the inverse (reciprocal) of the distance between locations $i$ and $j$ ($w_{ij} = 1/d_{ij}$)
  - Other functions also used
    - inverse of squared distance ($w_{ij} = 1/d_{ij}^2$), or
    - negative exponential ($w_{ij} = e^{-d}$ or $w_{ij} = e^{-d^2}$)
Measure of Spatial Autocorrelation
Global Measures and Local Measures

• Global Measures
  – A single value which applies to the entire data set
    • The same pattern or process occurs over the entire geographic area
    • An average for the entire area

• Local Measures
  – A value calculated for each observation unit
    • Different patterns or processes may occur in different parts of the region
    • A unique number for each location

• Global measures usually can be decomposed into a combination of local measures
Global Measures and Local Measures

• Global Measures
  – Join Count
  – Moran’s I

• Local Measures
  – Local Moran’s I
Join (or Joint or Joins) Count Statistic

- 60 for Rook Case
- 110 for Queen Case
Join Count: Test Statistic

Test Statistic given by:  \( Z = \frac{\text{Observed} - \text{Expected}}{\text{SD of Expected}} \)

\( \text{Expected} = \) random pattern generated by tossing a coin in each cell.

Expected given by:

\[
\begin{align*}
E(J_{BB}) &= kp_B^2 \\
E(J_{WW}) &= kp_W^2 \\
E(J_{BW}) &= 2kp_Bp_W
\end{align*}
\]

Standard Deviation of Expected (standard error) given by:

\[
\begin{align*}
E(s_{BB}) &= \sqrt{kp_B^2 + 2mp_B^3 - (k + 2m)p_B^4} \\
E(s_{WW}) &= \sqrt{kp_W^2 + 2mp_W^3 - (k + 2m)p_W^4} \\
E(s_{BW}) &= \sqrt{2(k + m)p_Bp_W - 4(k + 2m)p_B^2p_W^2}
\end{align*}
\]

Where: \( k \) is the total number of joins (neighbors)

\( p_B \) is the expected proportion Black, if random

\( p_W \) is the expected proportion White

\( m \) is calculated from \( k \) according to:

\[
m = \frac{1}{2} \sum_{i=1}^{n} k_i(k_i - 1)
\]
Gore/Bush Presidential Election 2000

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jbb</td>
<td>60</td>
</tr>
<tr>
<td>Jgg</td>
<td>21</td>
</tr>
<tr>
<td>Jbg</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
</tr>
</tbody>
</table>
### Join Count Statistic for Gore/Bush 2000 by State

<table>
<thead>
<tr>
<th>candidates</th>
<th>probability</th>
<th>Actual</th>
<th>Expected</th>
<th>Stan Dev</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bush</td>
<td>0.49885</td>
<td>60</td>
<td>27.125</td>
<td>8.667</td>
<td>3.7930</td>
</tr>
<tr>
<td>Gore</td>
<td>0.50115</td>
<td>21</td>
<td>27.375</td>
<td>8.704</td>
<td>-0.7325</td>
</tr>
<tr>
<td>Jbg</td>
<td>28</td>
<td>54.500</td>
<td></td>
<td>5.220</td>
<td>-5.0763</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>109</strong></td>
<td><strong>109.000</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The expected number of joins is calculated based on the proportion of votes each received in the election (for Bush = 109\*0.499\*0.499 = 27.125)
- There are far more Bush/Bush joins (actual = 60) than would be expected (27)
  - Positive autocorrelation
- There are far fewer Bush/Gore joins (actual = 28) than would be expected (54)
  - Positive autocorrelation
- No strong clustering evidence for Gore (actual = 21 slightly less than 27.375)
Moran’s I

- The most common measure of Spatial Autocorrelation
- Use for points or polygons
  - Join Count statistic only for polygons
- Use for a continuous variable (any value)
  - Join Count statistic only for binary variable (1,0)

Patrick Alfred Pierce Moran (1917-1988)
Formula for Moran’s I

\[
I = \frac{N \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}) \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

- Where:
  - \( N \) is the number of observations (points or polygons)
  - \( \bar{x} \) is the mean of the variable
  - \( x_i \) is the variable value at a particular location
  - \( x_j \) is the variable value at another location
  - \( w_{ij} \) is a weight indexing location of \( i \) relative to \( j \)
Moran’s $I$ and Correlation Coefficient

- **Correlation Coefficient** [-1, 1]
  - Relationship between **two** different variables
- **Moran’s I** [-1, 1]
  - Spatial autocorrelation and often involves **one** (spatially indexed) variable only
  - Correlation between observations of a spatial variable at location $X$ and “spatial lag” of $X$ formed by averaging all the observation at neighbors of $X$
Spatial auto-correlation

Correlation Coefficient

Note the similarity of the numerator (top) to the measures of spatial association discussed earlier if we view Yi as being the Xi for the neighboring polygon (see next slide)

Source: Ron Briggs of UT Dallas
\[
\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) / n
\]

\[
\sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}} \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}
\]

Yi is the Xi for the neighboring polygon

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x}) / \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

\[
\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

Moran’s I

Source: Ron Briggs of UT Dallas
Moran Scatter Plots

We can draw a scatter diagram between these two variables (in standardized form): $X$ and $\text{lag-X}$ (or $W_X$)

The slope of this regression line is Moran’s I
Moran Scatter Plots

Moran's I: 0.57328
Moran Scatterplot: Example

Moran's I: 0.633954
Statistical Significance Tests for Moran’s I

- Based on the normal frequency distribution with

\[ Z = \frac{I - E(I)}{S_{error(I)}} \]

Where:  
- \( I \) is the calculated value for Moran’s I from the sample
- \( E(I) \) is the expected value if random
- \( S \) is the standard error

- Statistical significance test
  - Monte Carlo test, as we did for spatial pattern analysis
  - Permutation test
    - Non-parametric
    - Data-driven, no assumption of the data
    - Implemented in GeoDa
Test Statistic for Normal Frequency Distribution

\[ Z = \frac{\text{observed Moran's I} - \text{expected Moran's I}}{\text{standard error of Moran's I}} \]

Reject null if \( Z > 1.96 \) (or \( Z < -1.96 \))

--- less than a 5% chance that, in the population, there is no spatial autocorrelation
--- 95% confident that spatial autocorrelation exists

Null Hypothesis: no spatial autocorrelation
*Moran’s I = 0

Alternative Hypothesis: spatial autocorrelation exists
*Moran’s I > 0

Reject Null Hypothesis if \( Z \) test statistic > 1.96 (or < -1.96)

*technically \(-1/(n-1)\)
Local Measures of Spatial Autocorrelation
Local Indicators of Spatial Association (LISA)

- **Local** versions of *Moran’s I, and the Getis-Ord G statistic*
- Moran’s I is most commonly used, and the local version is often called Anselin’s LISA, or just LISA

**See:**
Luc Anselin 1995 *Local Indicators of Spatial Association-LISA* Geographical Analysis 27: 93-115
Local Indicators of Spatial Association (LISA)

- The statistic is calculated for each areal unit in the data
- For each polygon, the index is calculated based on neighboring polygons with which it shares a border
- A measure is available for each polygon, these can be mapped to indicate how spatial autocorrelation varies over the study region
- Each index has an associated test statistic, we can also map which of the polygons has a statistically significant relationship with its neighbors, and show type of relationship
Summary

• Spatial autocorrelation of areal data
• Spatial weight matrix
• Measures of spatial autocorrelation
  • Global Measure
    – Moran’s I
• Local
  • LISA: Moran’s I
• Significance test
Diagnostic of Spatial Dependence

• Moran’s I is commonly used to diagnose if spatial dependence exists
  • For regression
    – calculate the residuals
      map the residuals: do you see any spatial patterns?
    – Calculate Moran’s I for the residuals: is it statistically significant?
Consequences of Ignoring Spatial Autocorrelation

- Autocorrelation/dependence -> information redundancy
  - Correlation coefficients and coefficients of determination appear bigger than they really are
    - You think the relationship is stronger than it really is
    - the variables in nearby areas affect each other
  - Standard errors appear smaller than they really are
    - exaggerated precision
    - You think your predictions are better than they really are since standard errors measure predictive accuracy
Consequences of Ignoring Spatial Autocorrelation

- Conventional statistical methods may not be reliable
  - E.g., for regression case, the assumption of *independence* is often required
  - *Spatial regression* is needed
Spatial Regression Methods

• Spatial Econometrics Approaches (available in GeoDa)
  – Lag model
  – Error model

• Other methods:
  • Spatial autoregressive model
  • Bayesian hierarchical models
• End of this topic