Spatial Analysis II – Point Pattern Analysis & Spatial Autocorrelation

Lex Comber
University of Leicester
ajc36@le.ac.uk
Aims

• Point pattern analysis
• Sampling
  – Test: Variance Mean Ratio
• Practical Pt1
• Spatial Autocorrelation
  – Joint count statistics
• Practical Pt 2
  – ArcGIS: to develop the mapping
Preamble

• Looked at 2 data formats
  – Vector: good for representing features; contains multiple attributes, lines, points and areas
  – Raster: good for data where geometry not relevant, regularly spaced data, no explicit information about features (implicit), continuous field of values

• The focus of this session is Point Patterns and the information that can be derived from point values
Point pattern analysis

- Example: a grid with values
  - It could be anything
    - house prices
    - obesity index
    - pollen counts
- What we are interested in is identifying any clustering in the data
- Is this clustered?
- Or is it random?
Point pattern analysis

• Example: a grid with *point* values
  – It could be anything
    • house prices
    • obesity index
    • pollen counts
• What we are interested in is identifying any clustering in the data
• Is this clustered?
• Or is it random?
Point pattern analysis

- Spatial patterns in data
  - Inform us about fundamental relationships in space or about possible causes for the observed patterns
- So by analysing the locations of incidences of \(<x\>\)...
  (You can pick your favourite disease, illness or public health topic at this point)
- ...we are often concerned with testing for
  - the occurrence of clusters
  - OR
  - the extent of randomness in the data
- Question: why might \(<x\>\) occur in clusters?
Point pattern analysis

- You have some data
- You need to think about
  - What are the outcomes for different patterns?
  - What criteria will you use?
- If it is non-random
  - Is it uniform?
  - Is it clustered?
Point pattern analysis

- Determine some measure that indicates
  - Uniform structures
  - Clustered structures
- Best: one number – a test statistic
- Examine many random patterns
  - determine what their test statistic values are
- If a given pattern has a measure that is much different from the ones of the random patterns
- Then the measure is unlikely for a random pattern, conclude pattern is not random
Point pattern analysis

• Summary
  – We are interested in identifying whether point patterns are random
  – Need to compare their pattern against what we would expect
  – Use a test statistic to do this
  – Apply to many random patterns to get the range of random values
  – If point pattern is outside this range, it may be random
Sampling

• Test statistic that indicates “randomness” of pattern
• What makes a random pattern?
• Can you tell me which of these is random and which may be clustered?
Sampling

• One way to do this is to impose a sampling grid
• Normally (ie under randomness) you would expect
  – some empty cells
  – many with about the mean number of points, and
  – few cells with many points
• Uniform: most have about the average number of points
• Clustered: many empty cells and with many points
Sampling

- Uniform: most cells have about the average number of points
  - Deviation of the number of points per cell from the mean is small
  - I.e. Small variance in points per cell

- Clustered: many empty cells and with many points
  - Deviation of the number of points per cell from the mean is large
  - Large variance in points per cell
Sampling

• Story so far
  – We have an idea of what we are looking for
  – Calculate
    • average number of points per cell
    • variance from that mean

• This describes a test statistic that will determine whether the observed pattern is different from expected
Test Statistic

• Variance Mean Ration (VMR)
  • Ratio of the variance of “the number of points per cell” divided by the mean number of points per cell.
    – Variable Y: number of points per cell
    – VMR = VAR(Y)/MEAN(Y)

• Tests on random data
  – VMR < 1 indicates uniformity
  – VAR(Y)=0 perfectly uniform
  – VMR > 1 indicates clustering
  – VMR ≈ 1 pretty random
Test Statistic

• Apply to the examples
  – very simple case
  – but good illustration

VMR = 0

VMR = 0.833

VMR = 1.33
Sampling / Test Statistic

• Sampling or **Quadrat Analysis**
  – A grid of square cells of equal size is overlaid on top of a x-y map of incidents
    • Can you see the GIS here?
  – Then you can count the number of data points in each grid cell.
  – In a random pattern the mean number of points in a cell will be roughly equal to the variance of the number of points per cell
  – This variance is measured statistically using the Variance Mean Ratio (VMR)
  – Quadrat size is crucial to the success of identifying spatial patterns and it has been suggested that the ‘optimal’ quadrat size contains 2 points per quadrat or cell
Sampling / Test Statistic

• Need to think about whether the test statistic result is significant
• Consider its sampling distribution
  – ie: the probability of the values of that test statistic in a situation in which the null hypothesis is true (i.e. under the null hypothesis)
• Its distribution is chi-squared with (n-1) degrees of freedom
• The test statistic here is
  – N x VMR (where N = number of grid cells)
Sampling / Test Statistic

• $p(t)$ – the probability of some random variable $X$ taking the value $t$

• Distribution shows which values of $X$ as likely or unlikely
Sampling / Test Statistic

• Shaded area is proportional to the probability that a value as large or larger than the observed VMR or could occur under the null hypothesis.
Test statistic and sampling distribution

Shaded area is proportional to the probability that a value as large or larger than the observed VMR or could occur under the null hypothesis.
Practical Pt1

- Excel spreadsheet of disease incidents
- Comprehensive instructions
- You should try to produce a short report (200-300 words) as part of this practical
<table>
<thead>
<tr>
<th>Easting</th>
<th>Northing</th>
<th>Col-Code</th>
<th>Row-Code</th>
<th>Cell-Code</th>
<th>Cell-Code</th>
<th>No. Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.726</td>
<td>0.702</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>0.137</td>
<td>0.299</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>0.899</td>
<td>0.856</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.835</td>
<td>0.433</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>0.135</td>
<td>0.999</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>0.812</td>
<td>0.418</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>0.019</td>
<td>0.806</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.835</td>
<td>0.861</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.858</td>
<td>0.844</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>0.882</td>
<td>0.834</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.491</td>
<td>0.575</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>0.954</td>
<td>0.924</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.113</td>
<td>0.500</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>0.490</td>
<td>0.181</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>0.011</td>
<td>0.152</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.797</td>
<td>0.721</td>
<td>4</td>
<td>3</td>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>0.828</td>
<td>0.840</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>0.619</td>
<td>0.243</td>
<td>3</td>
<td>1</td>
<td>19</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>0.344</td>
<td>0.115</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0.156</td>
<td>0.229</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.184</td>
<td>0.910</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>0.489</td>
<td>0.424</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>0.882</td>
<td>0.024</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.851</td>
<td>0.881</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>0.559</td>
<td>0.063</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>0.973</td>
<td>0.123</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>0.098</td>
<td>0.200</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>0.886</td>
<td>0.843</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>0.381</td>
<td>0.774</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>0.123</td>
<td>0.213</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.154</td>
<td>0.263</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0.121</td>
<td>0.260</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0.125</td>
<td>0.223</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0.107</td>
<td>0.221</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.170</td>
<td>0.200</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0.419</td>
<td>0.159</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0.481</td>
<td>0.125</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0.129</td>
<td>0.254</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0.996</td>
<td>0.061</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.446</td>
<td>0.543</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>0.017</td>
<td>0.705</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>0.185</td>
<td>0.198</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.640</td>
<td>0.856</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>0.630</td>
<td>0.303</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>0.738</td>
<td>0.766</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>0.188</td>
<td>0.993</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>0.718</td>
<td>0.635</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>0.881</td>
<td>0.310</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>0.452</td>
<td>0.798</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>14</td>
<td>1</td>
</tr>
</tbody>
</table>

**Summary Statistics**

- Number of Cells: 33
- Mean: 3.13
- Variance: 5.86
- VMR: 1.88

**Test Statistic:** 28.1

**p-value:** 0.021

**Disease Case Locations**

![Disease Case Locations](image)

**Study Area:** A

**Disease Case Analysis**
<table>
<thead>
<tr>
<th>Easting</th>
<th>Northing</th>
<th>Col-Code</th>
<th>Row-Code</th>
<th>Cell-Code</th>
<th>No. Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.839</td>
<td>0.954</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>0.900</td>
<td>0.906</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.722</td>
<td>0.782</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>0.636</td>
<td>0.034</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.095</td>
<td>0.331</td>
<td>1</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.942</td>
<td>0.093</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>0.214</td>
<td>0.075</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>0.252</td>
<td>0.803</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>0.656</td>
<td>0.888</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>0.075</td>
<td>0.871</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.688</td>
<td>0.776</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.068</td>
<td>0.125</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>0.200</td>
<td>0.142</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>0.800</td>
<td>0.176</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>0.622</td>
<td>0.154</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>0.741</td>
<td>0.522</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>0.387</td>
<td>0.336</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>0.238</td>
<td>0.491</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>0.190</td>
<td>0.923</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.332</td>
<td>0.304</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>0.631</td>
<td>0.135</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>0.727</td>
<td>0.940</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>0.033</td>
<td>0.224</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>0.948</td>
<td>0.867</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>0.093</td>
<td>0.078</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>0.953</td>
<td>0.397</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.565</td>
<td>0.050</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0.007</td>
<td>0.399</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.999</td>
<td>0.207</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>0.213</td>
<td>0.011</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>0.431</td>
<td>0.110</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0.075</td>
<td>0.656</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>0.602</td>
<td>0.519</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.492</td>
<td>0.186</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0.017</td>
<td>0.187</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.447</td>
<td>0.819</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>0.741</td>
<td>0.673</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>0.122</td>
<td>0.709</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0.999</td>
<td>0.043</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.514</td>
<td>0.856</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>0.660</td>
<td>0.918</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.131</td>
<td>0.804</td>
<td>1</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>0.958</td>
<td>0.273</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0.732</td>
<td>0.529</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
</tbody>
</table>

**Test Statistic:** 16.2

**p-value:** 0.336
Practical Pt 1

• What can you identify statistically about the randomness of the disease?
• Are the results the same for both study sites?
Summary Part 1

• Two components to Point Pattern Analysis
• 1) Find an indicator for the property of interest or the variable to be tested
  – Then apply test statistic, \( t \)
• 2) Determine the value of \( t \) under the null hypothesis
  – Then compare expected \( t \) with observed \( t \)
• If different from expected, reject null hypothesis
  – That is, the variation cannot be explained by randomness
• Sampling or quadrats provide a way of summarizing the data
• BUT the results will vary depending on the quadrat size and number
• Are there any questions?
  – I have given a fair bit of statistical information
  – But I hope you can see the links to the GIS analyses that you have already done
Spatial Auto-correlation

• Aim
  – understand autocorrelation
  – use and interpret Join Count statistics
Spatial Autocorrelation

• Autocorrelation in areal data
  – When value of a variable at one location is dependent on the value at another location nearby

• Explore whether there are local patterns of correlation in the data that might be hidden if we only investigate relationships between variables using linear regression (ie with space)

• Join Count tests for the absence of spatial autocorrelation
Spatial Autocorrelation

• Why not regression? Why spatial?
  – if simple regression is used only as a first guess, it can have serious drawbacks:
    • poor for local prediction – can produce too much global smoothing
    • if a lot of variables are included the regression coefficients, to preserve some spatial detail, the regression coefficients are highly correlated and the estimates unstable
    • local effects in one region will influence the fit of the regression everywhere
    • Variables may not be independent therefore
Spatial Autocorrelation

• Need to test for the *absence* of spatial autocorrelation

• Join-Count Statistic
  – Used to check for spatial autocorrelation
  – Explores pattern of contiguity of +ve and –ve classifications
  – Then asks if these are arranged in a pattern
Spatial Autocorrelation

- Consider a grid of raster data
- Some threshold value determined
- Each grid cell allocated into two groups based on whether below or above that value
  - Count the Joins
  - Negative to Negative
  - Positive to Positive
  - Positive to Negative
  - If threshold = 153

\[
\begin{array}{ccc}
N & N & P \\
N & P & P \\
N & N & P \\
\end{array}
\]

NN = 0, PP = 1, PN = 3

So effectively count the joins – thus Join Count statistic
Spatial Autocorrelation

- In this case the choice of threshold is key

\[ T = 155 \]
Spatial Autocorrelation

• Testing the null hypothesis
  – Normal distribution approximation
  – Selecting confidence level
  – If resultant values are greater than we would expect under a null hypothesis
    • reject hypothesis
    • conclude a pattern is present in residuals
  – Two-tailed test
Spatial Autocorrelation

• Interpretation of Joint Count Statistics
  – What does the result tell you?
  – Can you determine any pattern in the cells?
  – If so would it be related to something?
  – **Mid-level threshold** - tests whether ‘general’ clustering occurs
  – **Higher threshold** – tests ‘high-end’ clustering takes place
Spatial Autocorrelation

• Need to test significance
• If 8 white and 8 black cells could be due to
  – Free sampling: Each grid cell has a chance of 50% of being Black or White
    • Independent of the colour of the other cells
    • eg by drawing a balls from a hat with replacement
  – Unfree sampling:
    • eg by drawing a balls from a hat without replacement
  – Test for significance using Z statistic
    • More info on Z scores at
      • http://edndoc.esri.com/arcobjects/9.2/net/shared/geoprocessing/Spatial_Statistics_toolbox/what_is_a_z_score_qst_.htm
Practical Pt 2

• In ArcGIS / ArcMAP
  – Test for spatial autocorrelation using ‘Getis-Ord Gi*’
  – Other tests available
    • GWR, Moran’s I
  – Explore the influence of different sampling grids on the results
    • Regular grids
    • Census areas as a ‘grid’
References

• Rogerson, P.A., 2006. Statistical Methods of Geography. 2nd ed. Sage. London. Chap. 2.6, 10.1, 10.2