Decoding Health Data with Spatial Statistics

Flora Vale
Jenora D’Acosta
Wait a minute...
Wait a minute...

Where is Lauren??
Wait a minute...
Where is Lauren??
DAY 1
What are Spatial Statistics?
Measuring Geographic Distributions
Analyzing Patterns
Multivariate Analysis

DAY 2
Subjectivity of Maps
Mapping Clusters
Space Time Pattern Mining
Modeling Spatial Relationships
Are you using **Data** or **Information**?
Spreadsheets

Data or Information?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Maps

Data or Information?
When you look at a spreadsheet...
You ask for more

• Mean
• Standard Deviations
• Min and Max
• ...

1
2
3
4
5
Same goes for maps!
We can do more
What are Spatial Statistics?
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
Spatial Statistics Tools

Analyzing Patterns
  - Average Nearest Neighbor
  - High/Low Clustering (Getis-Ord General G)
  - Incremental Spatial Autocorrelation
  - Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
  - Spatial Autocorrelation (Morans I)

Mapping Clusters
  - Cluster and Outlier Analysis (Anselin Local Morans I)
  - Grouping Analysis
  - Hot Spot Analysis (Getis-Ord Gi*)
  - Optimized Hot Spot Analysis
  - Similarity Search

Measuring Geographic Distributions
  - Central Feature
  - Directional Distribution (Standard Deviational Ellipse)
  - Linear Directional Mean
  - Mean Center
  - Median Center
  - Standard Distance

Modeling Spatial Relationships
  - Exploratory Regression
  - Generate Network Spatial Weights
  - Generate Spatial Weights Matrix
  - Geographically Weighted Regression
  - Ordinary Least Squares
Spatial Statistics Tools

Analyzing Patterns
- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Incremental Spatial Autocorrelation
- Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
- Spatial Autocorrelation (Morans I)

Mapping Clusters
- Cluster and Outlier Analysis (Anselin Local Morans I)
- Grouping Analysis
- Hot Spot Analysis (Getis-Ord Gi*)
- Optimized Hot Spot Analysis
- Similarity Search

Measuring Geographic Distributions
- Central Feature
- Directional Distribution (Standard Deviation Ellipse)
- Linear Directional Mean
- Mean Center
- Median Center
- Standard Distance

Modeling Spatial Relationships
- Exploratory Regression
- Generate Network Spatial Weights
- Generate Spatial Weights Matrix
- Geographically Weighted Regression
- Ordinary Least Squares
python script!
### Remove Locations Outside Boundary FC ###

```python
featureLayer = "ClippedPointFC"
DM.MakeFeatureLayer(tempFC, featureLayer)
if self.boundaryFC:
    msg = ARCPRY.GetIDMessage(84454)
    ARCPRY.SetProgressor("default", msg)
    DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                             self.boundaryFC, ",#",
                             "NEW_SELECTION")
    DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                             ",#", ",#",
                             "NEW_SELECTION")
    DM.DeleteFeatures(featureLayer)
else:
    if additionalZeroDistScale == "ALL":
        msg = ARCPRY.GetIDMessage(84455)
        ARCPRY.SetProgressor("default", msg)
        DM.SelectLayerByAttribute(featureLayer, "NEW_SELECTION",
                                   ",Join_Count" = 0")
        DM.DeleteFeatures(featureLayer)
    else:
        distance = additionalZeroDistScale * fish.quadLength
        distanceStr = self.ssd0.distanceInfo.linearUnitString(distance,
                                                              convert = True)
        nativeStr = self.ssd0.distanceInfo.printDistance(distance)
        msg = "Removing cells further than %s from input points..." % nativeStr
        ARCPRY.AddMessage(msg % nativeStr)
        DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                  self.ssd0.inputFC, distanceStr,
                                  "NEW_SELECTION")
        DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                  ",#", ",#",
                                  "NEW_SELECTION")
        DM.DeleteFeatures(featureLayer)
DM.Delete(featureLayer)
del self.ssd0
ARCPRY.env.extent = oldExtent
```
Measuring Geographic Distributions
Descriptive Statistics
Mean
Median
Standard Deviation
Central Feature

identifies the most centrally located feature in a point, line, or polygon feature class
central feature
Mean Center

identifies the geographic center (or the center of concentration) for a set of features
mean center

mean = (17,15)
Median Center identifies the location that minimizes overall Euclidean distance to the features in a dataset.
X: 25, 24, 22, 18, 14, 14, 13, 12, 9
Y: 24, 23, 18, 16, 14, 12, 12, 12, 8

median = (14, 14)

median center
Mean vs Median?
mean = (33,28)
X: 74 25 24 22 18 14 14 13 12 2
Y: 38 24 23 18 16 14 12 12 12 8

median = (16,15)

mean = (33,28)
mean = (33, 28)

median = (16, 15)
Linear Directional Mean identifies the mean direction, length, and geographic center for a set of lines.
Standard Distance

measures the degree to which features are concentrated or dispersed around the geometric mean center
Directional Distribution
(Standard Deviational Ellipse)

creates standard deviational ellipses to summarize the spatial characteristics of geographic features: central tendency, dispersion, and directional trends
mean center
Analyzing Patterns
Global Inferential Statistics
Is there a PATTERN?
Clustered

Dispersed
Complete Spatial RANDOMNESS
z-scores and p-values

<table>
<thead>
<tr>
<th>z-scores</th>
<th>...-2.58</th>
<th>-1.96</th>
<th>-1.65</th>
<th>0</th>
<th>1.65</th>
<th>1.96</th>
<th>2.58...</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>...-0.01</td>
<td>-0.05</td>
<td>-0.10</td>
<td>0</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01...</td>
</tr>
</tbody>
</table>
z-scores and p-values

<table>
<thead>
<tr>
<th>z-scores</th>
<th>...-2.58</th>
<th>-1.96</th>
<th>-1.65</th>
<th>0</th>
<th>1.65</th>
<th>1.96</th>
<th>2.58...</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-values</td>
<td>...0.01</td>
<td>0.05</td>
<td>0.10</td>
<td>0</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01...</td>
</tr>
</tbody>
</table>
How intense is the clustering?
How intense is the clustering?

Compared to what?
How intense is the clustering?

Compared to where?
How intense is the clustering?
Compared to when?
An Example

Is the spatial segregation of the rich and the poor increasing or decreasing in New York?
An Example

1969

1985

2002

5.21

4.26

2.40
Average Nearest Neighbor

calculates a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature
average distance = expected average distance
ANN ratio = 

observed

expected
ANN ratio =

\[
\frac{\text{observed}}{\text{expected}} < 1
\]
ANN ratio = \frac{\text{observed}}{\text{expected}}

Clustered

\leq 1 \leq \text{Dispersed}
Spatial Autocorrelation
(Moran’s I)

measures spatial autocorrelation based on feature locations and attribute values using the Global Moran's I statistic
Are distances and values correlated?

Clustered

Dispersed
Incremental Spatial Autocorrelation

measures spatial autocorrelation for a series of distances and optionally creates a line graph of those distances and their corresponding z-scores
Spatial Autocorrelation by Distance
Spatial Autocorrelation by Distance

Distance (meters)

Z-score

60 80 100 120 140 160 180
Spatial Autocorrelation by Distance

- Fixed Distance Band = 70.3 meters
High/Low Clustering
(Getis-Ord General G)

measures the concentration of high or low values for a given study area
What type of clustering is present in the data?

High value clusters

Low value clusters
Multi-Distance Spatial Cluster Analysis (Ripleys K Function) determines whether features, or the values associated with features, exhibit statistically significant clustering or dispersion over a range of distances.
Spatial Clustering by Distance

Distance (meters) vs. \( L(d) \)
Spatial Clustering by Distance

- statistically significant clustering
- statistically significant dispersion

Distance (meters):
- 60
- 80
- 100
- 120
- 140
- 160
- 180

$L(d)$:
- 1
- 2
- 3
- 4
- 5
dispersed
clustered
Grouping Analysis

groups features based on feature attributes and optional spatial/temporal constraints
K Means
K Means

2 groups
K Means

2 groups

3 groups
K Means

2 groups

3 groups

4 groups
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
MATH!
interpret results through box plots
Similarity Search identifies which candidate features are most similar or most dissimilar to one or more input features based on feature attributes.
potential store locations
potential store locations

high performing store
Rank by how similar they are to based on:
- PopDensity
- AvIncome
- DistToCompetition

<table>
<thead>
<tr>
<th>LocID</th>
<th>PopDensity</th>
<th>AvIncome</th>
<th>DistToCompetition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Input Feature(s) to Match

Candidate Features

Attributes of Interest

<table>
<thead>
<tr>
<th>PopDensity</th>
<th>AvIncome</th>
<th>DistToCompetition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Input Feature(s) to Match

Candidate Features

Attributes of Interest

<table>
<thead>
<tr>
<th>PopDensity</th>
<th>AvIncome</th>
<th>DistToCompetition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 Match Methods
3 Match Methods

Attribute Values
3 Match Methods

Attribute Values

Ranked Attribute Values
3 Match Methods

Attribute Values

Ranked Attribute Values

Attribute Profiles
Attribute Values
Attribute Values

Z-transform:

\[
\frac{(x - \bar{x})}{SD}
\]
Attribute Values

- Population = 14,159
- % Uninsured = .26
- Distance (km) = 535.89
Attribute Values

standardize attributes

Population = -.7932
% Uninsured = 3.8462
Distance (km) = .6433
Attribute Values

- standardize attributes
- subtract candidate from target
- square differences
- sum squares
Ranked Attribute Values
Ranked Attribute Values

rank
attributes
Ranked Attribute Values

rank attributes

- 9.5
- 8.8
- 8.3
- 4.1
- 2.7
- 0.2
Ranked Attribute Values

rank attributes

- 9.5 → 6
- 8.8 → 5
- 8.3 → 4
- 4.1 → 3
- 2.7 → 2
- 0.2 → 1
Ranked Attribute Values

1. Rank attributes
2. Subtract candidate from target
3. Square differences
4. Sum squares

- sum squares
Attribute Profiles
Attribute Profiles

standardize attributes

cosine similarity index

\[
\text{cosine similarity index} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \sqrt{\sum_{i=1}^{n}(B_i)^2}}
\]

* Must have at least 2 attributes of interest
Dengue Fever Risk in Kenya
questions?
Want to learn more???

esriurl.com/spatialstats