Simple Ways to Do More with Your Data Using Spatial Statistics

Lauren Bennett
Jenora D’Acosta
Flora Vale

esriurl.com/spatialstats
What are Spatial Statistics?
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
proximity

area

orientation

coincidence connectivity

length
direction
Spreadsheets

Data or Information?

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

Data or Information?
When you look at a spreadsheet...

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>
You ask for more

1 2 3 4 5

• Mean
• Standard Deviations
• Min and Max
• …
Same goes for maps!
We can do more
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 (Global Moran's I)

Mapping Clusters
- Cluster and Outlier Analysis (Anselin Local Moran's I)
- Grouping Analysis
- Hot Spot Analysis (Getis-Ord Gi*)
- Optimized Hot Spot Analysis
- Optimized Outlier 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 (GWR)
- Ordinary Least Squares (OLS)
python script!
#### Remove Locations Outside Boundary FC ####

featureLayer = "ClippedPointFC"
DM.MakeFeatureLayer(tempFC, featureLayer)
if self.boundaryFC:
    msg = ARCPY.GetIDMessage(84454)
    ARCPY.SetProgressor("default", msg)
    DM.SelectLayerByLocation(featureLayer, "INTERSECT", self.boundaryFC, ",", "NEW_SELECTION")
    DM.SelectLayerByLocation(featureLayer, "INTERSECT", ",", ",", "SWITCH_SELECTION")
    DM.DeleteFeatures(featureLayer)
else:
    if additionalZeroDistScale == "ALL":
        msg = ARCPY.GetIDMessage(84455)
        ARCPY.SetProgressor("default", msg)
        DM.SelectLayerByAttribute(featureLayer, "NEW_SELECTION", ",Join_Count" = 0)
        DM.DeleteFeatures(featureLayer)
    else:
        distance = additionalZeroDistScale * fish.quadLength
        distanceStr = self.sssd.distanceInfo.linearUnitString(distance, convert = True)
        nativeStr = self.sssd.distanceInfo.printDistance(distance)
        msg = "Removing cells further than %s from input points...
        ARCPY.AddMessage(msg % nativeStr)
        DM.SelectLayerByLocation(featureLayer, "INTERSECT", self.sssd.inputFC, distanceStr, "NEW_SELECTION")
        DM.SelectLayerByLocation(featureLayer, "INTERSECT", ",", ",", "SWITCH_SELECTION")
        DM.DeleteFeatures(featureLayer)

DM.Delete(featureLayer)
de lýSSSD

ARCPY.env.extent = oldExtent
Measuring Geographic Distributions
Central Feature

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

identifies the geographic center (or the center of concentration) for a set of features
mean = (17, 15)
Median Center

identifies the location that minimizes overall Euclidean distance to the features in a dataset
median center

median = (14,14)
Mean vs Median?
\[ \text{mean} = (33, 28) \]
median = (16,15)

mean = (33,28)
median = (16,15)

mean = (33,28)
Demo
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
Demo
Analyzing Patterns
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
Demo
Average Nearest Neighbor

calculates a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature.
ANN ratio = \frac{\text{observed}}{\text{expected}}
ANN ratio = 

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

\begin{align*}
\text{Clustered} & \leq 1 \leq \text{Dispersed}
\end{align*}
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)

$L(d)$
dispersed
clustered
Mapping Clusters
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
<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>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rank by how similar they are to based on:
- PopDensity
- AvIncome
- DistToCompetition
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

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

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

1. Standardize attributes
2. Subtract candidate from target
3. Square differences
4. 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 6: 9.5
- Rank 5: 8.8
- Rank 4: 8.3
- Rank 3: 4.1
- Rank 2: 2.7
- Rank 1: 0.2
Ranked Attribute Values

1. rank attributes
2. subtract candidate from target
3. square differences
4. sum squares
Attribute Profiles
Attribute Profiles

- Standardize attributes
- Cosine similarity index

Graphs showing data trends.
Dengue Fever Risk in Kenya
Grouping Analysis

groups features based on feature attributes and optional spatial/temporal constraints
Grouping Analysis

with spatial constraints
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Crime in Chicago

- Median Income
- HS Dropout Rate
- Unemployment
- Crime Count
Grouping Analysis

with no spatial constraints
K Means
K Means

2 groups
K Means

2 groups

3 groups
K Means

2 groups

3 groups

4 groups
Demo
QUESTIONS?

Lauren Bennett  
Jenora D’Acosta  
Flora Vale

esriurl.com/spatialstats