Section 4

Video: Hot Spot and Outlier Analysis

Print

Time	Caption
0:00	♪ [music] ♪
0:11	The Hot Spot Analysis tool uses the Getis-Ord Gi* statistic
0:17	to identify statistically significant clusters of high and low values.
0:22	In other words, it finds places where values that
0:25	are very different from the average, either really high or really low,
0:29	cluster together spatially in a nonrandom way.
0:33	Let's use this set of polygons to illustrate
0:36	how a hot spot analysis works.
0:39	First, let's get clear on some terms.
0:42	Each polygon in this dataset is called a feature.
0:46	With the hot spot analysis, we're looking for clusters
0:49	of high and low values.
0:51	So each feature must have an associated numeric value.
0:55	The value can be something like a count, a rate, an average,
0:58	or any other numeric measure.
1:01	Values can be attributes associated with your features,
1:04	or you can aggregate incident data to produce a count value.
1:10	Next, every feature has what we call a neighborhood,
1:14	which is made up of its surrounding features
1:17	and includes the feature itself.
1:20	Finally, the study area refers to all
1:23	of the features in our dataset combined.
1:27	When we run the hot spot analysis, we're asking,
1:30	"Is this neighborhood significantly different from the study area?"
1:35	If the neighborhood's value is found to be
1:37	significantly higher than the study area,
1:40	the neighborhood's parent feature is marked as a hot spot.
1:44	So each feature's neighborhood is compared to the study area,
1:48	and the feature is assigned a probability
1:50	designating whether or not it belongs to a cluster.
1:54	The Hot Spot Analysis tool returns three levels of confidence.
1:59	We can be 90%, 95%, or 99% confident that a feature

2:03	belongs to a nonrandom cluster of high values, a hot spot,
2:08	or to a nonrandom cluster of low values, a cold spot.
2:11	And the colors in the resulting map correspond
2:14	directly to those confidence levels.
2:17	The Cluster and Outlier Analysis tool uses
2:20	the Anselin Local Moran's I statistic to identify
2:24	both value clusters and local outliers.
2:28	Similar to a hot spot analysis, a cluster and outlier analysis
2:32	finds places where high and low values cluster together,
2:35	and it also identifies features with values significantly different
2:40	from their surroundings, like a low value surrounded by high values,
2:44	or a high value surrounded by low values.
2:47	Just like with the hot spot analysis,
2:48	we have the feature, the neighborhood, and the study area.
2:53	But in a Cluster and Outlier Analysis,
2:55	the feature does not belong to its own neighborhood.
2:58	This allows for two different comparisons.
3:02	Is the neighborhood significantly different from the study area?
3:05	And is the feature significantly different from the study area?
3:09	If both the feature and the neighborhood have values
3:13	very different from the study area,
3:15	the feature is considered significant.
3:17	There are four possible types of significance.
3:21	If the feature and the neighborhood have similar values,
3:24	the feature is considered to belong to a cluster.
3:28	So a feature with a high value surrounded by a neighborhood
3:31	with a high value is marked as a high-high cluster,
3:35	while a feature with a low value surrounded by a neighborhood
3:38	with a low value is marked as a low-low cluster.
3:42	If the feature and the neighborhood have dissimilar values,
3:46	the feature is considered an outlier.
3:48	Outliers are features with values
3:51	very different from their surroundings.
3:53	So a feature with a high value surrounded by neighborhood
3:56	with a low value is marked as a high-low outlier.
3:59	And a feature with a low value surrounded by a neighborhood
4:03	with a high value is marked as a low-high outlier.

4:07	The resulting map shows the clusters
4:09	in the lighter shades of pink and blue
4:11	and the outliers in the brighter shades of red and blue.
4:15	Both of these statistical clustering methods help us identify
4:19	meaningful patterns in our spatial data by analyzing
4:22	features in the context of their immediate surroundings.
4:26	As Tobler's first law of geography states,
4:29	we expect things that are near each other to be related.
4:33	So an important part of these statistical cluster analysis methods
4:37	is defining the distance at which we expect the phenomena
4:41	that we are studying to be related.
4:44	This is called conceptualizing spatial relationships
4:48	or defining what it means to be a neighbor.
4:50	And there are many different ways to do this.
4:53	One of the more common methods is to use a fixed distance band.
4:58	With a fixed distance band, you specify a distance
5:01	of spatial influence, and all features within that distance
5:05	are considered to be spatially related or neighbors.
5:09	This is the method used in the Optimized Hot Spot
5:12	and Outlier Analysis tools, which interrogate your data to provide
5:16	a default distance at which clustering is most pronounced.
5:21	Another approach is to use contiguity.
5:24	With this method, anything that shares a border with a feature
5:28	will be considered its neighbor.
5:30	The k-nearest neighbors approach allows neighborhood sizes to vary
5:35	by including the feature's closest neighbors
5:38	no matter how far they actually are.
5:40	In this illustration, we've specified four nearest neighbors,
5:44	which for some features are a lot closer than others.
5:47	But despite the distance, those nearest features are still
5:51	the ones most likely to be related.
5:53	The Network Spatial Weights method is similar to a fixed distance band,
5:58	except that it measures distance along a network.
6:02	So you could specify everything within a 10-minute drive time
6:05	or a 5-mile driving radius, which is particularly useful
6:09	for capturing human behavior.
6:12	No matter which conceptualization of spatial relationships you choose,

6:16	it's extremely important that you consider both the subject matter
6:20	and your analysis question to thoughtfully define
6:23	what it means to be a neighbor.
6:25	♪ [music] ♪