Module 08 Practical

Kyle Lago

HLTH 661

20658108

02-July-2023

**PART A**

1.

A screenshot of a computer

Description automatically generated

2.

A screenshot of a computer

Description automatically generated

3.

A screenshot of a computer

Description automatically generated

4.

A screenshot of a computer

Description automatically generated

5. When comparing the prediction maps for the spline and inverse distance weighting methods, we can see a few differences. In our spline prediction map, the shades areas appear larger than the ones we see in the inverse distance weighted prediction map. This is likely due to the smoothing that occurs when using the spline algorithm that fits a curve through the data points. This smoothing is not a part of inverse distance weighting, so you will not see areas between data points being filled in.

Another reason could be the influence of distant points in inverse distance weighting. In this method, nearby points of data are weighted more heavily than distant points. The influence of these distant points may be reduced since they are further away from our interpolation point, and thus we may have smaller shaded areas.

We also used a power of 2 in our IDW method. As we increase power, we decrease the influence of distant data points which then can lead to smaller shaded areas. Using a power of 1 may have resulted in a prediction map that looked more like our spline prediction map.

6. The three methods we’ve used for spatial data interpolation all have their strengths and weaknesses.

Thiessen polygon method

This is the simplest method of the three we’ve used and it allows us to quickly breakdown data within a map. When we see the individual polygons, we know that inside each of them is the exact same value of the data point that is also contained within it. However, this also is one of its limitations. This method assumes that there is no spatial variation within the polygons.

An example would be the air quality of a region where polluting factory is located. Residents downwind from the factory on a given day could have differing air quality than those upwind. These differences might not be accounted for with Thiessen polygon interpolation if the residents all reside within the same polygon. Thiessen polygons only account for distance and would not consider other variations in geography that could be important factors.

Spline method

The spline method creates prediction maps with smoothed surfaces that pass through our data points. They are visually appealing and would likely be familiar to viewers as many weather radar maps use this sort of mapping to track inclement weather. They are not limited by artificial boundaries like the ones produced with Thiessen polygons. They also allow for variation around data points, unlike what we see within Thiessen polygons. It provides more reliable predictions as well.

However, this method does not do a good job of interpolating in areas where the terrain differs considerably and is best used in flat or gently varying surfaces. It is also sensitive to outliers, which can cause over-smoothing in the direction of these outliers. It also is not able to predict beyond the range of our observed data accurately or reliably.

Inverse Distance Weighting method

This method is simpler computationally than the spline method, as we are simply calculating the weighted average of nearby data points based on their inverse distances to where we are interpolating the data. It assigns greater weights to nearby data points, and thus is less susceptible to the influence of distant points. This makes intuitive sense if we consider the first law of geography, where we assume auto-correlation between nearby data points.

However, this method is quite sensitive to the power that is chosen when running the analysis. A power difference of one can change the prediction map quite considerably. If someone does not select the power correctly it can result in a prediction map with either over smoothing or under smoothing. This method is also very sensitive to outliers.

Conclusion

It is important to consider what data you are analyzing and what level of accuracy would suffice when choosing these methods. Thiessen and Inverse Distance Weighting are great for quick and simple approximations of the data. Spline is good for more complex spatial patterns. Spline and Inverse Distance Weighting are more visually appealing than Thiessen polygons (at least, in my opinion!).

**PART B**

1. We need a spatial reference system when working with data points on a map to maintain consistency and accuracy when assessing geospatial data. When we choose a spatial reference system and we let other knows which one we’ve used, they can comfortably use our data to reproduce similar results. This allows for reliable data sharing and analysis.

This is important in the sciences as it allows individuals to reproduce findings from other individuals which helps to ensure scientific integrity. If everyone used different referencing systems, it would be impossible to share data in a meaningful way. My XY coordinates might place a pin on the map somewhere completely different from another person, even though we are trying to reference the same physical location. Another good way to think of this is if you wanted to play a game with your friend.

If you both have different “rules” for the game, it would be difficult for you to play it together. By sharing the same set of “rules” it allows for mutual understanding on how to play together. Having the same set of rules may even allow you to play this game with people beyond your friend group, some who might not even speak the same language as you.

2.

A screenshot of a computer

Description automatically generated

3.

A screenshot of a computer

Description automatically generated

4.

A screenshot of a computer

Description automatically generated

5.

A screenshot of a computer

Description automatically generated

6.

A screenshot of a computer

Description automatically generated

7. There are significant differences in the amount of smoothing and the covered shaded areas between each map.

8.

Method 1 (power 1/neighbor 10)

PRO: Including more neighbors allows us to create a smoother surface and have more continuity between neighboring points. It provides a clean map and allows us to specify trends across region.

CON: Adding more neighbors may result in over smoothing, where regions between points have interpolation that is not accurate.

Method 2 (power 1/neighbor 5)

PRO: Including less neighbors results in less averaged variations between points. This combats against the effect of over smoothing that we seen in Method 1.

CON: There is less spatial continuity in this method, and we can see more abrupt changes. It can produce under smoothing that can imply the spread of disease differs significantly at these lines, when it in fact may not.

Method 3 (power 2/neighbor 10)

PRO: This provides the cleanest map and the most smoothing. Increasing the power also increases the influence of nearby points relative to distant ones.

CON: This method is much more sensitive to outliers and distance. Points nearest to our interpolated values exert much stronger influence than to those further away. Extreme outliers can also skew our shaded areas significantly depending on their locations.

9.

Power affects the influence of neighboring data points on our interpolated values. When we increase the power, we are giving more weight to data points that are nearer to our interpolated values than those that are further away. We can see this when we compare Method 2 and Method 3. The dark red region increase in size around our XY coordinates on the map.

Neighbors can affect the smoothing that we see in our surfaces. When we include more neighbors, this introduces more values that contribute to the average and thus more smoothing. We can see this when we compare Method 1 and Method 2. These both use the same power but reduce the number of neighbors in Method 2 to 5. This results in a surface with more abrupt and jagged “boundaries” than the ones we see in Method 1 that are smoother.

Adjusting the power and number of neighbors allows one to fine tune their interpolation based on whatever their desired outcomes are in geospatial analysis. Increasing the power will result in larger shaded areas near our points of interest, while increasing neighbors will result in more smoothing across areas.

10.

A screenshot of a computer

Description automatically generated

11.

A screenshot of a computer

Description automatically generated

12.

A screenshot of a computer

Description automatically generated

13. A screenshot of a computer

Description automatically generated

14.

A screenshot of a computer

Description automatically generated

15.

A screenshot of a table

Description automatically generated with low confidence

16.

A screenshot of a computer screen

Description automatically generated with low confidence

As we can see from the statistics generated for our prediction columns, Method 2 had highest standard deviation and thus the most variance among the methods we chose for interpolation. I would consider this method the least reliable for the interpolation of this data due to the low number of neighbours.

In this practical we are covering the entire country of Uganda and all its districts. The AIDS data is collected by counting cases from people reporting to hospital. There may be individuals in rural regions that do not have access to the hospital and thus they are not included in the dataset. Increasing the number of neighbors in our interpolation helps address this uneven distribution of data across the country. I believe that it makes sense to increase the number of neighbors for interpolation using this data.

Method 1 and 3 are better for this reason, but they differ in their power. Increasing our power makes it so points nearby are weighted more heavily. Since the points on our map are focused on the geographic centers of the region, it might make more sense to reduce the power of our model. This will provide us a more generalized representation of AIDS cases across the country and not be so specific to the centroids we have plotted.

For these reasons I believe Method 1 is the superior interpolation method for our goals.

17. I believe given the nature of AIDS that utilizing a semivariogram and spatial weight matrix could benefit the program. AIDS is a sexually transmitted disease, and individuals are more likely to have sex within close proximity to them than those further away. This is how it spreads from one isolated case to around an entire community. Employing a semivariogram would allows us to identify AIDS clusters in the country that differ from one another spatially. Within those clusters, we could then employ a spatial weight matrix to break them down even further, which allows us to identify hotpots and explore why they have come to exist.

A semivariogram would allow us to look at the spatial autocorrelation of AIDS cases across Africa. We could see the range of influence which would tell us when AIDS cases become less like each other. This is important because the reasoning for spread of AIDS in one region may be different than the spread in another. We could find that our semivariogram has a high sill, which would mean there is great variability in AIDS cases across the country. This could help us single out large AIDS clusters in the different regions of Uganda that are unrelated to each other.

Once these clusters are identified, we can begin exploring them in more detail. Two clusters may differ as to why AIDS is proliferating within them. Researchers could collect relevant data and then implement a spatial weight matrix. Calculating Moran’s I would allow researchers to identify hotspots within these clusters statistically, and then they can begin to explore why these hotspots exist.

Maybe one hotspot within an AIDS cluster does not provide free condoms to the community. Maybe another does but does not educate individuals on how to use them effectively. They may look similar in overall case rates, but they would differ as to why and this is important for intervention. Utilizing a semivariogram and spatial weight matrix would allow researchers to feel confident that they’ve identified regions where AIDS is proliferating. Without these methods, it would be difficult to identify them empirically. These decisions are very important when it comes to implementing public health interventions as funding is often scarce.