Geography 464- Assignment 1

Point Pattern Analysis

Due Date: Tuesday Feb 8th, 2011, 5pm
(with complete Chi-Squared calculation sheet)
Total points available: 100 (5 for each question)

Objective 1: Reviewing the concepts and procedures of calculating Mean Center, Standard Distance, as well as Standard Deviation Ellipse;

Q1.a) Please describe the geographic location of three types of crimes and their relative location to the mean center of all-crimes shapefile. (You can use the DC_Boundary.shp to help you)
All three mean centers are found near the mean center of all crimes, one is above, one is below to the right and the other is below to the left. This would make sense that the mean center of all crimes falls in the middle of specific crimes.

Q1.b) Please relate to some information of District of Columbia (you might also want to Google map or Google earth it) and describe the neighborhood where a certain type of crime happens more frequently (i.e. residential area, central business district, or other)
When comparing the points of all crimes to an aerial image of DC it is possible to locate the river in the data and then find out that the majority of the dots fall where there are larger buildings like hotels, this making me believe it to be downtown in the city where there might be more tourists for example, more so than residential areas. However there are a few more crimes in the residential areas than say the US Capital or Parks for example (obviously the river).

Q1.c) The standard distance of All Crime is 4891.5793, the standard distance of Robbery is 5022.45802, the standard distance of Burglary is 4871.1251, and the standard distance of Vehicle Crime is 4859.54594.

Q1.d) The crime with the least dispersion is __Vehicle Crime____, while the crime with the most dispersion is _Robbery_. (Only consider the three specific crime types)

Q1.e) What is the general orientation of each crime distribution? Are all three crimes in the same orientations?
All three crimes appear to have the same size ellipse, not exactly however, with similar orientation. All three are longer in the NW/SE direction than they are NE/SW.

Q1.f) Which ellipse seems the largest and smallest to you? Can you predict its relationship to the dispersion of a crime? (You can borrow some of the conclusion you derived in some previous questions)

Robbery appears to have the largest ellipse, also verified this with the info tool and comparing std. X,Y distances. This follows what was previously discovered with robbery having the most dispersion with the largest standard distance.

Objective 2: **Using two methods of estimating point densities to describe the density of three different crimes in areas where robbery, burglary, and vehicle crime are concentrated;**

Q2.a) How would you describe the overall density of crimes in areas where burglary, robbery, and vehicle crimes are found? How would you describe the relationship between overall crime density and crime type?

Overall, the three categories are all have more density circles towards the same area as before where the ellipses were centered. Clustering is clearly visible but there are some outliers. Overall crime is very similar to the other 3 combines, this makes sense and it does have a slightly larger footprint but this is due to being able to combine all the categories when coming up with the densities.

Q2.b) Explain what the "nonuniformity problem" is? Do you think this problem appears in this step of carrying out point density?

The nonuniformity problem comes up when dealing with local densities due to the area and center points changing. I believe this does come up but the program has a way of filling in the blanks.

Q2.c) Does the map produced by a kernel density estimator look different from the point density map? If so, how did it happen even though you used the same search radius and output cell size?

The two maps are slightly different, the Kernal map seems to have a smaller footprint of densities. One reason they are different is a Kernal density is more sensitive to the radius, with more emphasis on weighted distances. Far events will most likely be left out and given a density of 0.

Q2.d) Would differences between these two maps lead you to draw different conclusions about the relationship between overall crime density and crime type, or is the difference mainly "cosmetic"? If you are going to report these results to an official at the police department, which map will you use and why?

From my perspective and use the differences seem cosmetic as they are very similar and depending on the use I suppose it may matter but a crime sense I can't imagine it a difference of a block allowing catastrophic events to occur. Based on that I would say it would make sense to use the kernel map as it focuses more on the events that occur really close together, and would allow patrol units to further focus their routes instead of using a broader footprint obtained previously.

Objective 3: Use density-based spatial statistics to test hypotheses regarding the spatial distribution of trees within specific age classes.

Q3.a) For **old trees** at a 35m scale (9 quadrat case), state your null and alternative hypotheses, and report your Chi-squared value, degrees for freedom, and P-value. Do you accept, or reject your null hypothesis? Do you feel that the statistical inference made by this analysis matches up with your perception of how the data "look"?
Explain.

The null hypothesis would state that there should be an expected value of 20/9 trees per quadrant, alternatively this value is different. The Chi-squared value was 25.56. The degrees freedom for this example would be the number of categories, 9 quadrants, minus 1, so 8. Using the chi-squared value and 8 df we come up with a p value of .001, which is significant to reject the null hypothesis and suggest clustering.

Q3.b) For **mature trees** at a 35m scale (9 quadrat case), report your Chi-squared value, degrees for freedom, and P-value. Which analyses that you have previously done (either descriptive or hypothesis testing) lead you to the same conclusion about the distribution of mature trees as this chi-squared analysis? Explain.

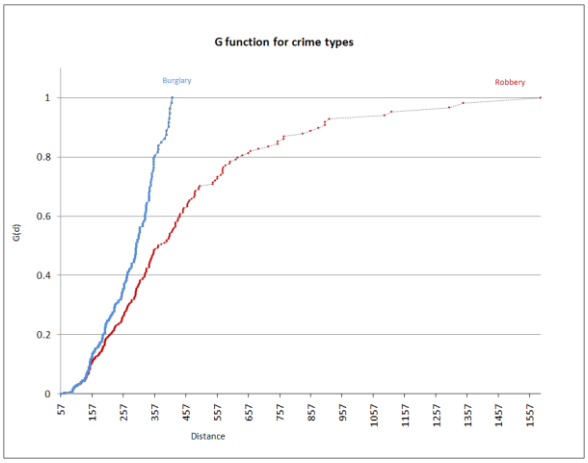
The chi-squared value for mature trees was 67.33, with 8 df. This gives a p-value of 0.000. Descriptive statistics based on observation can relate to this chi-squared analysis in that the results verify the almost linear clustering.

Q3.c) For **young trees** at a 35m scale (9 quadrat case) report your Chi-squared value, degrees for freedom, and P-value.
For young trees the chi-squared value was 33.47 with 8 df with a p value of 0.00005.

Name a quadrat count analysis you could perform to test the alternative hypothesis that old tree and young trees are not found together in the same quadrats? If the answer is not in your notes, you can find it in your textbook.
Assuming that there is clustering amongst the two with each other, you could test the hypothesis that they are not found together by combining the counts and you would see if the data is still clustered or more evenly distributed. A remaining cluster would lead to the fact that to are found in the same quadrats, while a more dispersed result would suggest less overlap.

Q4.a) For **robbery**, what is the z-score and ratio for the observed to expected NNDs, and what can you conclude about their spatial dispersion? What is the **null** hypothesis you are testing? At what range of z-scores would you reject the **null** hypothesis at the significance level of 0.05?
The z-score was -9.55 which suggests clustering. The null hypothesis would be that there is no clustering. We would reject the null at a z score of +- 1.96, for a significance level of 0.05; we therefor reject this and assume clustering is present.

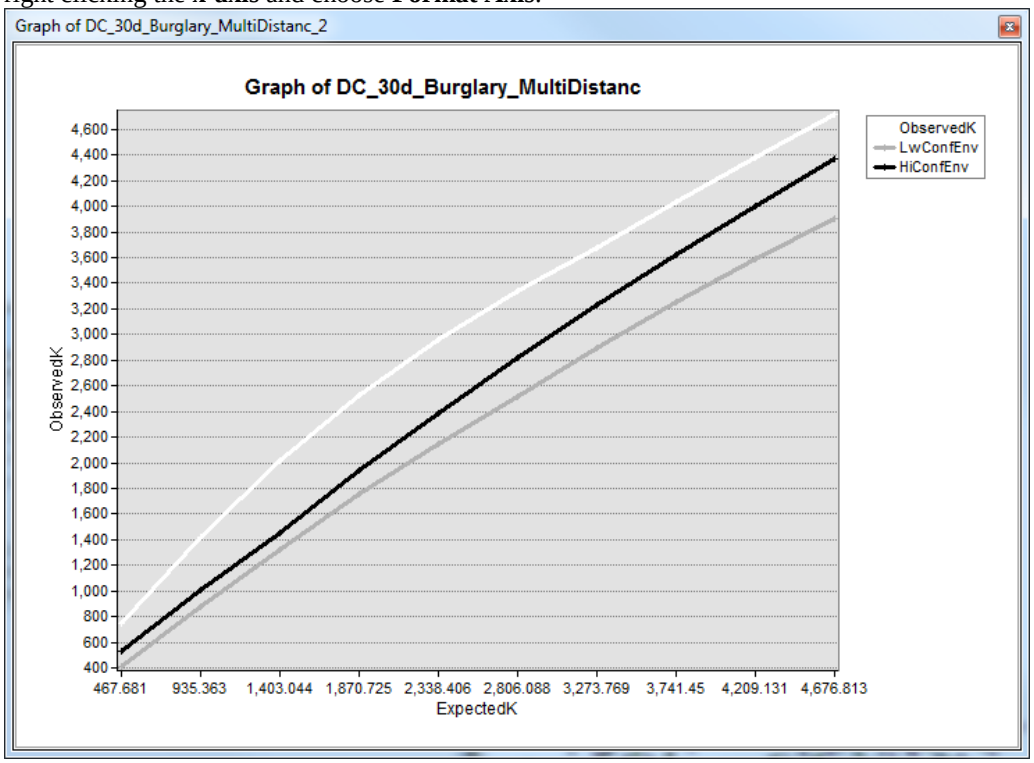
Q4.b) For **burglary**, what is the z-score and ratio for the observed to expected NNDs, and what can you conclude about their spatial dispersion?
With a z score of -7.67, we can also conclude there is a significant evidence of clustering but not as much as robbery.
Q4.c) For **vehicle crime**, what is the z-score and ratio for the observed to expected NNDs, and what can you conclude about their spatial dispersion?
With a z score of -19.71, we can also conclude there is a significant evidence of clustering, more than the previous 2.
Q4.d) Which crime has a larger proportion of its crimes with a nearest neighbor distance of 300m or less?
According to this graph, which crime appears to be more evenly dispersed?
Burglary has a larger proportion of its crimes with a NND of < 300m. According to the graph Robbery would seem to have a more dispersed pattern based on the events occurring over a larger NND range.



Q4.e) What are some limitations/shortcomings in using NND and the G function as a means to test for spatial clustering?

These two tests only use the nearest neighbor for each event, this can have a great effect when a NND is very small relative to other distances in the same study.

Q4.f) In your PPA_Spring10 dataset, you can find a file **Vehicle_NND.dbf**, please open it in Excel and create a G Function graph similar to the one above. (When you are creating the chart, remember to change you setting of Axis Type to **Data Axis** in **Axis Format** window. You can get this Axis Format window by right clicking the **x-axis** and choose **Format Axis**.



Q4.g) According to your graph, at what spatial scales (if any) does burglary appear to be significantly clustered, and at what scales (if any) do they appear to be significantly dispersed? Do you observe any distances at which the pattern is random?

According to the graph there is minimal clustering, with significant dispersion across the entire range represented with what is nearly a line. The slight curve tho does indicate some clustering from around 1000 to 3000 ExpectedK. Nothing appears to be random as I assume that would be a jagged line.

Q4.h) Comparing the two methods of G function and Ripley's K analysis. Do these two analyses produce different results? Please comment on the differences of these two methods?

When comparing vehicle crime using G and K functions there are some differences in the results. The G function is almost an inverse of what the K function produced. The difference is probably related to the fact that the K function is more dependent on the proximity of the occurrences which would obviously produce different results than the G function. I would imagine clustering may be more evident using Ripley's K analysis for areas where clustering cannot be determined by eye.