**Lab 2: Spatial statistics and probability distributions**

Geog 4300/6300-Shannon

Fall 2016

Value: 20 points

**Overview**: This lab covers two main topics: basic spatial statistics and probability distributions. Your answers to the lab questions should be typed in a separate Word document and turned in using the Assignment Dropbox on the ELC site.

**Spatial statistics:**

1. The lab folder on Google Drive has two files with crime data for Spokane, Washington. “Spokane\_crimes\_all” has data for every crime committed in from 2000-2015. “SpokanePrecinct\_2014\_crimecounts” has the number of crimes in 2014 broken down by type for each voting precinct in the city. Download these files to your computer as csv files.  
     
   ***Question 1 (3 points):*** *Use Rstudio or Excel to open the first data set, Spokane\_crimes\_all. Calculate the mean center for (1) all crimes, (2) all robberies (see the “Offense” column)*. *The “subset” command can be used to select just robberies. What are the coordinates of each? Interpret how any difference between them (e.g., direction and distance) is meaningful*

2. You can also use the spatstat package to create a ppp object for your robbery data. Plot out a density map, adjusting the kernel size to one you think best shows the data. See the lab script from week 6 for more information.

* Plot(density([object],[kernel size]))

Then create a quadrat count using quadratcount:

* quadratcount([object], [# of cols],[# of rows])

***Question 2 (3 points):***  *Combine the kernel density map and appropriately sized quadrat grid for robberies in Spokane during the study period, including it in your response. You’ll need to include “add=TRUE” for the quadratcount function to layer it, and possibly change the text color as we did in class. Briefly explain what this graph tells you about robbery “hot spots” within the city and why you chose the parameters you did (kernel density and quadrat size).*

1. You can calculate the equity of distribution of crimes using R. Load the “SpokanePrecinct\_2014\_crimecounts” file. Crimes are commonly reported as crimes/100,000, so you’ll need to calculate that rate. Do so by creating a new variable in the data, dividing the assaults by population and then multiplying by 100,000:  
     
   [new variable] <- [assault variable] / [population variable] \* 100000  
     
   Install and load the “ineq” package in R. To calculate the Gini coefficient for the distribution of these crimes, use the command “Gini()” for your new rate column.   
     
   To plot this distribution as a Lorenz curve, you need to type two commands. Assume that x is your variable of interest. These commands would read: (1) LC.x<-Lc(x) (2)plot(LC.x). (Note: LC.x is an arbitrary title. Any new variable name should work.) Plot the Lorenz curve for the Spokane precinct crime data.  
     
   ***Question 3 (2 points):*** *Include both the Gini coefficient for this crime data and your Lorenz curve plot. Interpret both of them—what do these tell you about the distribution of violent crime in Spokane? Be specific about how both the Gini coefficient and Lorenz curve support this conclusion.*

***Question 4 (3 points):*** *The residents of precinct 3104 are wondering how their assault rate compares to the rest of the city. The “SpokanePrecinct\_2014\_crimecounts” file shows both total crimes and assaults by precinct. Calculate the location quotient for assaults as a percent of crimes in this neighborhood compared to the city as a whole using either R or Excel. Explain how you did so and interpret the number you produced.*

**Probability:**

1. The number of persons in each camping party that reserves a campsite at Dawgsville National Park is assumed to be distributed as a Poisson distribution.  The maximum number of persons in a camping party is 8.  The mean number of persons in a camping party is 3.43.

***Question 5 (3 points):*** *Create a table that shows the probability of observing each possible count of persons in a camping party for each reservation (including no shows).*

The average score on the final exam in a statistics course over the last 10 years is 82%, with a standard deviation of 7.15.  These scores were normally distributed.  
  
***Question 6 (3 points):*** *Assuming that the distribution of scores is normally distributed, what is the probability that a randomly selected student will score more than 90% on the final exam?  What score on the final exam marks off the bottom 25% of scores?*

For the past 100 years, a stream close to Athens has been measured at a gauging station. The station measures the volume passing a point in a minute, so the values obtained are in units of cubic feet per minute (or CFM). A USGS scientist has crunched these numbers and determined that the results are normally distributed, with a mean of 35.26 CFM and a standard deviation of 4.61 CFM  
  
***Question 7 (3 points):*** *Using this information, calculate the z-score for each of the following levels, which could produce damaging levels of flooding: 40 CFM, 44 CFM, and 48 CFM. Using the table found on p. 303 of your text (or the pnorm function), what is the probability that any randomly selected observation will fall above each of these amounts?*

**Bonus!!**

For up to 2 additional points, create two maps using Leaflet. The first should show the locations of all robberies in 2009. The second should show the mean centers for all crimes and just robberies for the whole time period (the two points you found in Question 1), color coding the two points. Export each from RStudio as a screen shot or image and insert them into your Word document.