STA130 Homework #3: Comparison of Homicide Rates

(Due in tutorial at 2:10 pm sharp on Wednesday, March 9, 2016.)

NOTE: You may discuss this assignment with other students, but you must understand and type and create your solutions entirely by yourself. You can also ask your TA for assistance, either after a tutorial or by arranging to meet with them another time. And, don't forget the information at www.probability.ca/Rinfo.html about the "R" statistical software package.

INTRODUCTION:

One way to compare crime levels in different cities is by looking at the rate (per 100,000 population) of homicides (i.e., murders) on an annual (i.e., year-by-year) basis. In this assignment, we will compare the homicide rates of three large Canadian cities: Toronto, Montreal, and Winnipeg. The goal will be to determine which city is safest, and which city is least safe. Because the homicide rates vary year by year, in essentially a "random" way, statistical analysis is required.

Based on the above information, consider the following questions. Answer them as best as you can, making use of "R" as necessary. You should then prepare answers to hand in at your tutorial. Your answers should include explanations in complete English sentences, together with precise numerical answers where appropriate. You should also include, with explanation, all of the "R" commands you used.

1. Obtain the relevant data! Go to CANSIM Table 253-0004 at the Statistics Canada web page at: http://www5.statcan.gc.ca/cansim/pick-choisir?id=2530004

From there, go to the "Add/Remove data" tab, and under "Geography" add the Toronto and Montreal and Winnipeg options, and under "Time Frame" choose the year range from 1985 to 2014, and under "Homicides" be sure that the "Homicide rates per 100,000 population" option is checked. Then click the "Apply" button (in Step 5 at the bottom) to retrieve the data. You should see, among others, a row for Toronto homicide rates (beginning with 2.24 in 1985, then 1.28 in 1986, etc., up to 1.38 in 2014), and rows for the Montreal and Winnipeg rates too.

2. Read the data into R! Find a way to create three R vectors, one for each of the three cities. Each vector should be of length 30, and include the homicide rate for that city for each year 1985–2014. (This operation might be a bit "messy" – it is often tricky to get data from a web page into an R vector, so much so that it has a name: "data scraping". One method is to simply copy-and-paste the relevant rows into a file, and then fix them up as necessary (e.g. adding commas?) so they can be read into R. You could also try to use R's "read.table" method (with the "as.is" option set to TRUE), but that is tricky because the web page also includes lots of other items besides the table of data. Or you could try the "Download" tab at the top of the web page and see if those options help. Good luck!)

3. Output your three R vectors, and check carefully to confirm that you have indeed read the data correctly into R. (Include these data values in the answers you hand in, either as vectors of numbers or as graphs.)

4. For each of the three cities' annual homicide rates, do the following:

(a) Plot a graph of the homicide rate, as a function of the year.

(b) Compute the observed mean and variance and sd.

(c) Compute a 95% confidence interval for the true mean.

5. Plot a single graph of all three cities' homicide rates, as a function of the year, all on the same graph together, hopefully in different colours and/or with different symbols. (Hint: After plotting the first city's rates with "plot", try using the R command "points" for the remaining rates, so it adds them to the same plot; see also the "col" and "pch" options.)

6. For each of the three pairs of cities (i.e., for Toronto-Montreal, and Toronto-Winnipeg, and Montreal-Winnipeg), compute a 95% confidence interval for the difference between the true means of the two cities.

7. For each of the three pairs of cities, compute a P-value to test the null hypothesis that the true means of the two cities are equal.

8. Interpret the results of all of the above computations clearly and explicitly and completely, in complete English sentences. In particular, do your computations establish which city is the safest, and which city is the least safe, or don't they? Explain.

Hand in your answers and R commands and plots and printed output to your TA, right at the start of tutorial at 2:10 pm on Wednesday March 9.