Bryson Cook

ISYE6501, Spring 2018

Homework 3

**Question 5.1**

**Using crime data from http://www.statsci.org/data/general/uscrime.txt (description at http://www.statsci.org/data/general/uscrime.html), test to see whether there is an outlier in the last column (number of crimes per 100,000 people). Is the lowest-crime city an outlier? Is the highest-crime city an outlier? Are there others? Use the grubbs.test function in the outliers package in R.**

When sorting the data into ascending order and plotting, it does not appear like the lowest crime rate is an outlier though it appears there are outliers at the highest crime levels, though it is not certain particularly which. 1969 and 1993 are the highest crime rates, and are well above the next highest rate group, which begins at 1674 and includes the next two points. So one could also state that the top 5 results are outliers from the rest, as there is a large gap between points 42 and 45 (1272 vs 1555).

When applying the Grubbs test to the upper end, the P Values are as follows:

[,1] [,2]

[1,] 1993 0.07887486

[2,] 1969 0.02847821

[3,] 1674 0.17807968

[4,] 1635 0.11389923

[5,] 1555 0.10815574

[6,] 1272 1.00000000

[7,] NA NA

[8,] NA NA

If the Grubbs test determines that the point is an outlier that point is removed and the test is run on the next point until a P value of 1 is reached, signifying that the point is not an outlier. Here, the 6th point in question, 1272, was reached and determined to not be an outlier.

The process was then repeated for the lower range, with no points showing as an outlier, as the first point's pvalue equals 1, meaning no points were determined to be outliers and no data was removed.

[,1] [,2]

[1,] 342 1

[2,] NA NA

[3,] NA NA

[4,] NA NA

[5,] NA NA

[6,] NA NA

[7,] NA NA

[8,] NA NA

**Question 6.1**

**Describe a situation or problem from your job, everyday life, current events, etc., for which a Change Detection model would be appropriate. Applying the CUSUM technique, how would you choose the critical value and the threshold?**

**Question 6.2**

**1. Using July through October daily-high-temperature data for Atlanta for 1996 through 2015, use a CUSUM approach to identify when unofficial summer ends (i.e., when the weather starts cooling off) each year. You can get the data that you need from the file temps.txt or online, for example at http://www.iweathernet.com/atlanta-weather-records or https://www.wunderground.com/history/airport/KFTY/2015/7/1/CustomHistory.html . You can use R if you’d like, but it’s straightforward enough that an Excel spreadsheet can easily do the job too.**

1. Apply the CUSUM to each column of data, just need to pick the right amount of cooling. Based on the average of the years, we can determine the when the end of summer
2. Average all the columns together and apply CUSUM to the average.
   1. Assume S0 = 0
   2. Mu could be the mean of July temps
   3. Example of T is 5x the std dev of the summer temps
   4. Plot St over time and detect change when St >= T

**2. Use a CUSUM approach to make a judgment of whether Atlanta’s summer climate has gotten warmer in that time (and if so, when).**

Is summer getting longer?

Is summer getting hotter?