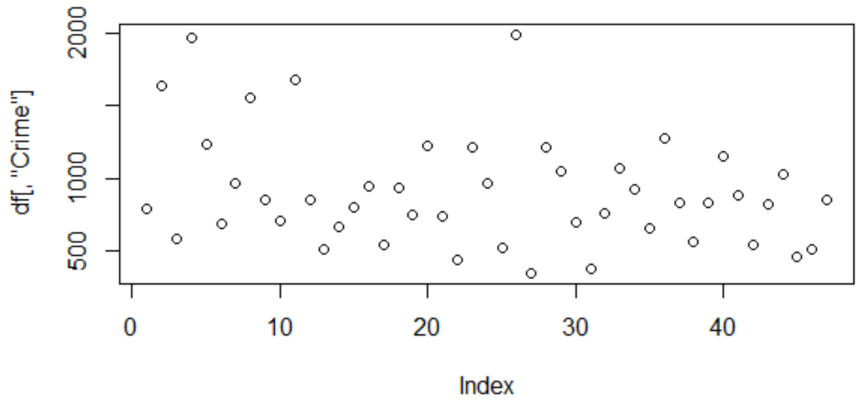
ISYE 6501 HW3

February 2021

1 Question 5.1 Detect Outlier

First I reviewed that there are 47 data points in the uscrime data set, which is a fairly small data set. So we can use grubbs.test function in R to check outliers. To have an idea about the data nature, I have plot a scatter plot for the Crime column. We can tell from the plot that the data are distributed pretty evenly within the range.



Then I'm applying the grubbs.test function to Crime column, and I'm testing both the largest and smallest data points.

- H0 hypothesis test: There is no outlier in the data.
- Ha hypothesis test: The tested point is an outlier.

The largest point which has a value of 1993 gives the test result of G = 2.81287, U = 0.82426, p-value = 0.07887, where p-value is greater than 0.05 that I can't reject H0, which means that the largest point 1993 is not an outlier.

The smallest point which has a value of 342 gives the test result of G = 1.45589, U = 0.95292, p-value = 1, where p-value is greater than 0.05 that I can't reject H0, which means that the smallest point 342 is not an outlier.

Since both the largest and smallest points are not outliers, I will conclude that there is no outlier in the Crime column.

1.1 My R code is:

```
library(outliers)

df <- read.table("data 5.1/uscrime.txt", sep = '\t', stringsAsFactors = FALSE, header = TRUE)

str(df[,"Crime"])
plot(df[,"Crime"])

# there are 47 data points, it's small sample, so we can use grubbs.test to check outliers

# test largest point
one_outlier_h <- grubbs.test(df[,"Crime"], type = 10, opposite = FALSE)

# test smallest point
one_outlier_l <- grubbs.test(df[,"Crime"], type = 10, opposite = TRUE)
```

2 Question 6.1

An example for change detection would be to review if a treatment is in fact effective for some disease/condition. The target value would be whatever the normal standard for the specific disease/condition, we can also use simulation to simulate a critical value. Threshold is normally 5 standard deviation but in reality it would truly depend on the disease/condition.

3 Question 6.2 CUSUM

3.1 Question 6.2 a

I'm using Excel to process this method. First, I use July and August temperatures to find the mean for each year, because I'm pretty confident that the July and August are considered as summer. Thus the mean for only July and August would be a good target value for CUSUM. Second, I used all 4 months temperature data to find the standard deviation for each year.

Then I use the CUSUM function $S_t = max\{0, S_{t-1} + (\mu - x_t - C)\}$ to get CUSUM value for each day in each year. There are many "false alert" before the actual "end of summer day", which are reasonable due to cooling turbulence or temporary weather change in the warm days. (In tab "a) Computing")

I pick the date right after the last 0 appeared in a year as the "end of summer" day. This is a more conservative approach since this date would for sure be a day not in summer no matter if there was any weather effect (example: temperature already cooled down but some warm turbulence hit the area and temperature got temporarily raised). I sorted the 20 dates and take the median as the estimated end of summer day for Atlanta, which is Sept 10th or Sept 11th (in tab "a) Sort Date"). Take these 2 days in mind and go back to the data, I found that for all 20 years, there were already some cool days before Sept 10/11 (positive number in "St year" column), which confirmed my estimation.

3.2 Question 6.2 b

For this question, I decide to use Sept 11th to complete my estimation. (in tab "b)") First I calculated the mean temperature for summer days for each year using the Sept 11th as end of summer day. Then calculate the mean and standard deviation over the 20 years. Now I have all values I need to complete CUSUM calculation, I plug them in the function to calculate increase trend $S_t = max\{0, S_{t-1} + (x_t - \mu - C)\}$. The result are plot in the following chart, we can see that the trend went up and down, it had a high peak during the year 2010 to 2013 but went back to 0 for 2014 and 2015. Thus I wouldn't say that there is clear evidence that the temperature went up over time.