

nb_week3

January 30, 2019

1 Homework Week 3

1.1 Question 5.1

Using crime data from the file `uscrime.txt` (<http://www.statsci.org/data/general/uscrime.txt>, description at <http://www.statsci.org/data/general/uscrime.html>), test to see whether there are any outliers in the last column (number of crimes per 100,000 people). Use the `grubbs.test` function in the `outliers` package in R.

```
In [4]: library(outliers)
```

```
In [ ]: data <- read.table("uscrime.txt", sep="\t", header=TRUE)
```

Testing for an outlier:

```
In [16]: grubbs.test(data$Crime, type=10, opposite=FALSE)
```

Grubbs test for one outlier

```
data: data$Crime
G = 2.81290, U = 0.82426, p-value = 0.07887
alternative hypothesis: highest value 1993 is an outlier
```

The value of 1993 has only a probability of 7.9% of being by chance. Let's remove the value and test again:

```
In [37]: data <- data[-match(c(1993), data$Crime),]
         grubbs.test(data$Crime, type=10, opposite=FALSE)
```

Grubbs test for one outlier

```
data: data$Crime
G = 3.06340, U = 0.78682, p-value = 0.02848
alternative hypothesis: highest value 1969 is an outlier
```

The value of 1969 has only a probability of 2.8% of being by chance. Let's remove the value and test again:

```
In [38]: data <- data[-match(c(1969), data$Crime),]  
        grubbs.test(data$Crime, type=10, opposite=FALSE)
```

Grubbs test for one outlier

```
data: data$Crime  
G = 2.56460, U = 0.84712, p-value = 0.1781  
alternative hypothesis: highest value 1674 is an outlier
```

The value of 1674 has a probability of 17.8% of being by chance. That's not an outlier!

1.2 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?

Since I am working in the aerospace industry, I often get confronted with the issue of process control. One example is the milling of highly complex parts from titanium. To detect wear out of the cutter, we could monitor the forces and power consumption of the machine to detect changes. To choose the critical values and the threshold we could use past data of failing cutters.

1.3 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.iweather.net.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.

The Calculations for Question 6.2 are in the attached Excel File.

The cumsum approach with values of 5 * standard deviation for T and 1/2 * standard deviation for C returns the end of summer as:

Year	End of Summer
1996	30. Sep
1997	16. Oct
1998	07. Oct
1999	06. Oct
2000	06. Oct
2001	30. Sep
2002	14. Oct

Year	End of Summer
2003	30. Sep
2004	11. Oct
2005	12. Oct
2006	14. Oct
2007	18. Oct
2008	19. Oct
2009	14. Oct
2010	01. Oct
2011	09. Oct
2012	07. Oct
2013	18. Oct
2014	28. Sep
2015	30. Sep

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).

If we apply the CUMSUM approach to the mean temperature of each summer, the result suggests that there is an increase in temperature in the year 2010. But since then the temperature drops again to the level of the previous years, we can not conclude that the summer climate got warmer in general.