Subject: ISYE6501

Assignment: HM2

Students (name, GT id):

×          Yuanting Fan, 904047984

×          Wenjia Hu, 904057780

×          Sen Yang, 904025995

***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?**

***Example 1:***

To monitor deforestation and afforestation in a region, the CUSUM technique may be applied to detect small monthly changes in forest areas, triggering an alert when the forest-to-land ratio falls below a specified threshold.

* **Critical value and threshold**: 0.5 times the standard deviation of the forest-to-land ratio over a given period. However, due to seasonality or random events the critical value will be adjusted based on the frequency and magnitude of forest changes in the area. The T value can be set as the region’s historical forest-to-land ratio from the past 5 years. Similarly, the threshold might be adjusted dynamically.

***Example 2:***

We designed an accelerator app for mobile game users to help them launch games faster. The number of users launching a game simultaneously varies at different times, and during high levels of concurrency, the accelerator may crash. Therefore, we want to monitor for any anomalies that indicate an increasing trend in crashes.

* **Critical value and threshold**: we set a low C value, specifically, we use 0.05% for the C value and 1% for the T value. This means that the minimum recorded change is a 0.05% variation in the crash rate. Any cumulative crash rate change of 1% signals a significant change.

***Example 3:***

Employee absences are common, but HR must monitor attendance closely to maintain control. There are two scenarios where detecting the shifts of employee’s attendance is required: whether the employee attendance situation has significantly declined so that interventions are needed, and after interventions, whether the attendance situation has improved so that the company should continue invest in those intervention measures.

* **Critical value and threshold:** In both scenarios, with a 90% average attendance and 5% standard deviation, HR sets a 9% critical value to detect a 10% change. Thresholds vary by company—those with higher salaries or lower performance may use smaller thresholds due to the higher cost of absenteeism.

***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**](http://www.iweathernet.com/atlanta-weather-records) **or** [**https://www.wunderground.com/history/airport/KFTY/2015/7/1/CustomHistory.html**](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.**

***Data overview:***

We are given a dataset of the daily maximum temperature of Atlanta from July to October over a period of 20 years. An initial look at the data shows that temperature changes occur gradually over this period, with the daily maximum temperatures fluctuating within a small range. This scenario is well-suited for applying a change detection method. By calculating the cumulative sum of temperature deviations from the target temperature (mean), we can monitor smaller shifts in the data and therefore identify temperature drops.

***Data analysis:***

We adopt the function: =Max (=0) to detect temperature drops in each year. As the logic of CUSUM is quite straightforward, the key to CUSUM analysis lies in determining a target value μ (usually the mean of the in-control situation, and in this case, the mean temperature), a critical value C (a constant used to reduce random effects and adjust sensitivity), and a threshold T (a value that determines how quickly a significant change can be identified).

**Step 1: What is the expected value for the mean μ?**  
The first step in CUSUM analysis is to choose a suitable μ. In general, μ represents the expected value when no significant temperature changes occur. In this case, it should be the average daily maximum temperature of summer in Atlanta. The question then becomes how to define the summer period in Atlanta.

In astronomy, summer refers to the period between the summer solstice and the autumnal equinox. In the Northern Hemisphere, the summer solstice falls on or around June 21, while the autumnal equinox occurs around September 22.[[1]](#footnote-1) In this assignment, we use astronomical summer and thus choose the July-September data to calculate the target for our analysis.

**Step 2: Determining C and T**

A review of research papers on CUSUM shows that in general C is set to half of the shift in standard units, i.e. C=/ 2, and T is usually chosen to be 4 or 5 times the standard deviation. [[2]](#footnote-2) Since the temperature changes are small shifts, we set T to 4 times the standard deviation to make our analysis more sensitive to temperature changes and detect them faster.

|  |  |
| --- | --- |
| Year | Summer end date |
| 1996 | 19-Sep |
| 1997 | 25-Sep |
| 1998 | 25-Sep |
| 1999 | 20-Sep |
| 2000 | 7-Sep |
| 2001 | 25-Sep |
| 2002 | 24-Sep |
| 2003 | 10-Sep |
| 2004 | 16-Sep |
| 2005 | 5-Oct |
| 2006 | 15-Sep |
| 2007 | 18-Sep |
| 2008 | 17-Sep |
| 2009 | 1-Sep |
| 2010 | 27-Sep |
| 2011 | 7-Sep |
| 2012 | 18-Sep |
| 2013 | 25-Sep |
| 2014 | 25-Sep |
| 2015 | 16-Sep |

This table shows the result we obtain after applying the CUSUM approach we designed above. As shown in the table, except for 2005, when the summer ended on Oct 5, Atlanta’s summer ended in September in all the other years, which is consistent with the astronomical definition and general expectations.

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

To identify whether the summer climate has become warmer from 1996 to 2015, we use the CUSUM approach again to detect whether there are significant changes in the average summer daily high temperatures and the total length of summer in Atlanta over the 20 years. We calculate the two variables based on the summer end dates we obtained in the previous question.

In both cases of the average summer temperatures and summer length, the critical value C and threshold T are set to half of and 4 times the corresponding standard deviations, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Summer end date | Average temp | Summer length |
| 1996 | 19-Sep | 88.23 | 81 |
| 1997 | 25-Sep | 86.06 | 87 |
| 1998 | 25-Sep | 87.22 | 87 |
| 1999 | 20-Sep | 88.75 | 82 |
| 2000 | 7-Sep | 89.98 | 69 |
| 2001 | 25-Sep | 85.42 | 87 |
| 2002 | 24-Sep | 88.3 | 86 |
| 2003 | 10-Sep | 85.66 | 72 |
| 2004 | 16-Sep | 85.26 | 78 |
| 2005 | 5-Oct | 86.44 | 97 |
| 2006 | 15-Sep | 88.55 | 77 |
| 2007 | 18-Sep | 89.78 | 80 |
| 2008 | 17-Sep | 87 | 79 |
| 2009 | 1-Sep | 86.85 | 63 |
| 2010 | 27-Sep | 90.97 | 89 |
| 2011 | 7-Sep | 91.82 | 69 |
| 2012 | 18-Sep | 89.45 | 80 |
| 2013 | 25-Sep | 84.72 | 87 |
| 2014 | 25-Sep | 86.78 | 87 |
| 2015 | 16-Sep | 88.33 | 78 |
| SD |  | 1.97023449 | 8.136176523 |
| average |  | 87.7785 | 80.75 |
| T=4\*SD |  | 7.880937962 | 32.54470609 |
| C=0.5\*SD |  | 0.985117245 | 4.068088262 |

We then obtain the cumulative sums of the average summer temperatures and summer length as the table shows below:

|  |  |
| --- | --- |
| CUSUM for average temp | CUSUM for summer length |
| 0 | 0 |
| 0 | 0 |
| 0 | 0 |
| 0 | 0 |
| 1.21638275 | 7.68191174 |
| 0 | 0 |
| 0 | 0 |
| 0 | 4.68191174 |
| 0 | 3.36382348 |
| 0 | 0 |
| 0 | 0 |
| 1.01638275 | 0 |
| 0 | 0 |
| 0 | 13.6819117 |
| 2.20638275 | 1.36382348 |
| 5.26276551 | 9.04573521 |
| 5.94914826 | 5.72764695 |
| 1.90553102 | 0 |
| 0 | 0 |
| 0 | 0 |

As shown in the two charts above, in both cases, the CUSUM algorithm does not detect a cumulative sum of over 4 times the standard deviations, indicating that there are no significant changes in either the average summer temperatures or summer length during the period. Therefore, we come to the conclusion that Atlanta’s summer climate has not become warmer between 1996 and 2015.

Extra:

我看了CUSUM的一些资料，说设置SIGMA的前提是假设数据正态分布

# Shapiro-Wilk 检验统计量 = 0.955, p 值 = 0.000

# 数据不服从正态分布

A graph with a red line

Description automatically generated

# 这个Q-Q图也不是正态分布的

首先，我抛弃之前看到的C,T要几个sigma的限制，CUSUM的本质是测试到什么时候超出了我的容忍度；我自己定义夏天温度是“7-8月”的平均温度，而夏天的结束是在9-10月份。我能容忍的最大降温幅度T是7-10月份的最高温度和最低温度的两倍（或者1.5倍），每天允许的降幅C是7-10月份的平均降温。所以我在测试的是8月份开始降温之后降到T的幅度就标志夏天结束了，因为累计降幅超过了我这四个月的最大降幅两倍（因为是累计降幅，所以不能是1倍）

# 测出来的夏天结束大部分是9月中下旬

1. National Weather Service. (n.d.). *Astronomical seasons*. National Oceanic and Atmospheric Administration. <https://www.weather.gov/twc/Astronomical_Seasons> [↑](#footnote-ref-1)
2. Alwan, L. C. (2000). *Statistical process analysis*. McGraw-Hill International Editions. [↑](#footnote-ref-2)