**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.*
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).*

Here are two approaches. Please note that there are other good approaches too, and a good solution doesn’t have to try all of the possibilities in here; they’re shown to help you learn, but they’re not necessary. It also might be that some people interpreted the question differently, and have a good solution to the question as they understood it – that’s perfectly fine too. All I really care about is that you’re learning how to use the CUSUM model.

The file solution 6.2.xlsx (sheet Part 1 Approach 1) shows the first approach. The “temps” sheet shows the raw data, and the “Part 1 Approach 1” sheet shows a CUSUM calculation. (Note: Column C is set to turn green starting at the first detected change, using conditional formatting.) For each day (July 1 through October 31), we take the average temperature over the 20-year period 1996-2015. To define , the average temperature before any change, we use the average of the average July temperatures (i.e., the average of: July 1’s average, July 2’s average, …, July 31’s average)[[1]](#footnote-1). That turns out to be 88.75, which is pretty reasonable compared to the July average temperature range from 87 to 90.4.

By changing the values in cells D1 and F1, you can test different values of C and T, some of which are shown in this chart:

|  |  |  |  |
| --- | --- | --- | --- |
| **C** | **T** | **Change detected** | **Avg temp on change date** |
| 5 | 25 | Sep 26 | 78.55 |
| 5 | 15 | Sep 24 | 80.4 |
| 5 | 5 | Sep 18 | 81.7 |
| 1 | 25 | Sep 9 | 85.05 |
| 1 | 15 | Sep 6 | 84.6 |
| 1 | 5 | Sep 1 | 85.25 |
| 0.44262 | 4.42625 | Aug 30 | 85.8 |

Table 1. Change detection for different values of C and T. The last row shows C as ½ standard deviation of the July data, and T as 5 standard deviations of the July data, which is sometimes used as a default.

My personal judgment (which you don’t have to agree with – there’s not a single “right” answer) is that the default values are too sensitive. There is a small change (from the 87-88 range to the 85 range) that starts on August 30, but that change seems too small to call the end of summer. Something closer to Sep 24 or 25 seems like a more-significant change. On the other hand, the graph below does seem to show a change around the end of August (close to Labor Day, the traditional end of summer).

Figure 2. Average (1996-2015) temperature day by day.

A second approach is shown in the file solution 6.2.xlsx (sheet “Part 1 Approach 2”). In this approach, we find the change point for each year individually, and then average them.

In the “CUSUM” sheet, each column B through U corresponds to a year. B12:B134, C12:C134, etc. are the values of St for 1996, 1997, etc. Rows 1-4 show the July average, max, min, and standard deviation each year. Rows 5 through 7 show the values of C, T, and mu each year. Rows 8 and 9 show the date a change is detected each year. Finally, cells V8 and V9 show the average day that change is detected.

As in the first approach, the key is finding the right values of C and T. Looking at the year-to-year data, there are certainly some days where the temperature is temporarily 10-15 degrees colder than average, but then goes back up – we don’t want to mistakenly identify that randomness as a change.

|  |  |  |  |
| --- | --- | --- | --- |
| **C** | **T** | **Avg date change detected** | **Avg temp on change date** |
| 0 | 35 | Aug 18 | 90.15 |
| 5 | 35 | Sep 19 | 82.4 |
| 10 | 35 | Oct 7 | 75.6 |
| 0 | 25 | Jul 22 | 89.45 |
| 5 | 25 | Sep 17 | 79.75 |
| 10 | 25 | Oct 4 | 77.85 |
| 0 | 15 | Jul 17 | 89.2 |
| 5 | 15 | Sep 2 | 85.25 |
| 10 | 15 | Sep 26 | 78.55 |

Table 2. End-of-summer estimates using approach 2.

As above, this is a judgment call, but my preference would be for C=5 and T=25 or 35, giving a date around Sep 17 or Sep 19. Figure 3 below shows the year-by-year data.

Figure 3. Year-by-year end-of-summer estimates.

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).*

Again, here are two approaches. Please note that there are other good approaches too, and a good solution doesn’t have to try all of the possibilities in here; they’re shown to help you learn, but they’re not necessary. It also might be that some people interpreted the question differently, and have a good solution to the question as they understood it – that’s perfectly fine too. All I really care about is that you’re learning how to use the CUSUM model.

# One approach is to take the yearly end-of-summer date estimate from Part 1 Approach 2, and see if there’s a change over time. In other words, “is summer lasting longer than it did before?“ This is shown in solution 6.2.xlsx (sheet “Part 2 Approach 1”). The values of C and T we choose make the difference between a “yes” or “no” answer.

|  |  |  |  |
| --- | --- | --- | --- |
| **C** | **T** | **Year change detected** | **Change detected in year** |
| 3 | 60 | Never | n/a |
| 3 | 45 | 2007 | Oct 13 |
| 3 | 30 | 2005 | Oct 9 |
| 3 | 15 | 1999 | Sep 22 |
| 7 | 60 | Never | n/a |
| 7 | 45 | Never | n/a |
| 7 | 30 | Never | n/a |
| 7 | 15 | 2007 | Oct 13 |
| 14 | 60 | Never | n/a |
| 14 | 45 | Never | n/a |
| 14 | 30 | Never | n/a |
| 14 | 15 | Never | n/a |

# Table 3. End-of-summer estimates.

As you can see, this is ambiguous. It’s not clear if summer is ending later. In fact, as Figure 6 shows below, if anything there was a blip in 2012-2013 where our Part 1 Approach 2 model shows summer ending significantly *earlier*.

# Figure 4. Year-by-year end-of-summer estimates (same as Figure 5).

A second approach is to answer “however long summer is, are those days getting hotter?” As shown in solution 6.2.xlsx (sheet “Part 2 Approach 2”, for each year we could take the end-of-summer estimate from Part 1 Approach 2, and calculate the average daily temperature through that date.

Figure 4 shows the year-by-year average summer temperature estimates. Visually, it seems difficult to identify a change point – 2010-2012 appears higher, but then the 2013-2015 drop back to the standard range. CUSUM similarly is ambiguous about whether/when there was a change; see Table 6. For more-sensitive values of C and T, change is detected in 2011 or 2012, but other values of C and T don’t detect a change, because the temperatures drop back starting in 2013.

# Figure 5. Year-by-year average summer temperature estimates.

|  |  |  |  |
| --- | --- | --- | --- |
| **C** | **T** | **Year change detected** | **Avg temp through day** |
| 0 | 5 | 2011 | 92.16 |
| 0 | 10 | Never | n/a |
| 0 | 15 | Never | n/a |
| 1 | 5 | 2012 | 91.85 |
| 1 | 10 | Never | n/a |
| 1 | 15 | Never | n/a |
| 2 | 5 | Never | n/a |
| 2 | 10 | Never | n/a |
| 2 | 15 | Never | n/a |

# Table 4. End-of-summer estimates using approach 2.

1. Of course, this is the same as the average over all 620 July observations (31 days each year for 20 years). [↑](#footnote-ref-1)