

## HW2 ISYE 6501 Summer 2018 OA

**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**

### How CUSUM works:

The goal in the CUSUM<sup>1</sup> (cumulative sum) approach is to detect change. Specifically, looking at the formula For detecting a change that is "higher" than normal:

First we calculate the

$$S(t) == \max(0, S(t-1) + x(t) - \mu - C)$$

and then ask: Is  $S(t) \Rightarrow T$

- The idea here is to calculate the cumulative sum  $S(t)$  and then comparing it (for each observation) against a threshold  $T$ 
  - $S(t)$  is the maximum of either 0 or
  - the addition of prior  $S(t-1)$  and the difference for the observed data point from its average ( $x(t) - \mu$ )
- the  $C$  parameter is a dampener to change the models sensitivity
  - higher  $C$  means the model is less sensitive, but may be delayed (or may never) detect the change
  - lower  $C$  means the model is more sensitive
- the threshold  $T$  is decided upon based on how costly it is to qualify that a change is detected
- the value  $C$  depends on how costly it is to be too sensitive to change.

### Goal of this exercise...

**To evaluate all the daily temperature values of the 4month data for 15 years, and see when temperature dramatically starts to drop off**

- Essentially this is the \*opposite\* of the above equation.
- Meaning, we are trying to measure the change has dropped below threshold:
- So , re-framing the equation:

$$\text{If } S(t) == \max(0, S(t-1) + \mu - x(t) - C)$$

Then is Is  $S(t) \Rightarrow T$  ?

---

<sup>1</sup> Extra reading for CUSUM here: <https://itl.nist.gov/div898/handbook/pmc/section3/pmc323.htm>

- extra reading talks about v masks here:

Steps followed in creating the model

- First I set up the excel sheet (google sheets) where I calculated the Average, Standard Deviation, Target and Dampener for EACH year..
  - Average and Stdev are built-in excel functions and self-explanatory
  - For the Target T, I used any change which is 2 standard deviations from the norm. I made this into a variable as seen below so I can change it and see the effect
  - For the dampener C, I started with 0 first and changed it as I saw the spurious values show up..
- I then created new columns just to calculate S(t) as per the above equation. (Col X in Figure 6.2.1)
- Also created a simple IF statement to see if S(t) > T. If yes, then I marked that as TRUE, otherwise I marked it FALSE (Column Y in Figure 6.2.1)
  - Initially, I set up the model with just S(1996) but then later, once this year was tuned I replicated this for S>T for all years.

Variables														
Target (T)	2 standard deviations from norm													
Dampener (C)	0 standard deviations from norm													
AVERAGE:	83.72	81.67	84.26	83.36	84.03	81.55	83.59	81.48	81.67	83.94	83.30			
STDDEV:	8.55	9.32	6.41	9.72	9.52	8.22	9.43	7.02	7.73	6.59	8.71			
Target (T)	17.10	18.64	12.82	19.45	19.04	16.45	18.85	14.04	15.45	13.18	17.42			
Dampener (C)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015			
1-Jul	98	86	91	84	89	84	90	73	82	90	85	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-14.28	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-13.28	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-6.28	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-5.28	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-9.28	0	FALSE
7-Jul	93	75	93	82	96	87	89	87	79	89	90	-9.28	0	FALSE
8-Jul	91	87	95	86	91	89	89	90	88	90	90	-7.28	0	FALSE
9-Jul	93	84	95	87	96	91	90	89	88	90	91	-9.28	0	FALSE
10-Jul	93	87	91	87	99	87	91	84	87	87	93	-9.28	0	FALSE
11-Jul	90	84	91	82	96	90	84	84	80	85	92	-6.28	0	FALSE
12-Jul	91	88	86	77	93	90	77	86	87	90	93	-7.28	0	FALSE
13-Jul	93	86	88	73	91	86	82	87	78	89	92	-9.28	0	FALSE
14-Jul	93	90	87	81	93	82	88	84	85	90	90	-9.28	0	FALSE

Figure 6.2.1 - Model Snapshot

### How I determined the value of C:

Started with C = 0. Immediately noticed in the chart, spurious S(t) values spiking and then dropping back down, as marked in the Orange Cells below.

Variables														
Target (T)	2 standard deviations from norm													
Dampener (C)	0 standard deviations from norm													
AVERAGE:	83.72	81.67	84.26	83.36	84.03	81.55	83.59	81.48	81.67	83.94	83.30			
STDDEV:	8.55	9.32	6.41	9.72	9.52	8.22	9.43	7.02	7.73	6.59	8.71			
Target (T)	17.10	18.64	12.82	19.45	19.04	16.45	18.85	14.04	15.45	13.18	17.42			
Dampener (C)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-14.28	0	FALSE
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-13.28	0	FALSE
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-13.28	0	FALSE
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-6.28	0	FALSE
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-5.28	0	FALSE
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-9.28	0	FALSE
7-Jul	93	75	93	82	96	87	89	87	79	89	90	-9.28	0	FALSE
8-Jul	91	87	95	86	91	89	89	90	88	90	90	-7.28	0	FALSE
9-Jul	93	84	95	87	96	91	90	89	88	90	91	-9.28	0	FALSE
10-Jul	93	87	91	87	99	87	91	84	87	87	93	-9.28	0	FALSE
11-Jul	90	84	91	82	96	90	84	84	80	85	92	-6.28	0	FALSE
12-Jul	91	88	86	77	93	90	77	86	87	90	93	-7.28	0	FALSE
13-Jul	93	86	88	73	91	86	82	87	78	89	92	-9.28	0	FALSE
14-Jul	93	90	87	81	93	82	88	84	85	90	90	-9.28	0	FALSE
15-Jul	82	91	91	81	93	82	91	86	86	86	89	1.72	1.71544715	FALSE
16-Jul	91	91	87	86	93	84	93	88	87	83	88	-7.28	0	FALSE
17-Jul	96	89	90	82	91	87	93	88	91	86	93	-12.28	0	FALSE
18-Jul	95	89	91	87	97	88	93	88	87	82	92	-11.28	0	FALSE
19-Jul	96	89	95	88	100	90	93	88	90	85	91	-12.28	0	FALSE
20-Jul	99	90	91	90	99	87	91	88	86	76	93	-15.28	0	FALSE
21-Jul	91	89	91	90	93	84	95	89	87	82	93	-7.28	0	FALSE
22-Jul	95	84	89	91	96	87	91	86	85	83	92	-11.28	0	FALSE
23-Jul	91	87	91	93	87	90	89	81	84	88	88	-7.28	0	FALSE
24-Jul	93	88	91	93	82	84	87	82	86	87	91	-9.28	0	FALSE
25-Jul	84	89	86	91	75	82	84	84	89	88	90	-0.28	0	FALSE
26-Jul	84	89	88	93	82	88	86	87	86	89	91	-0.28	0	FALSE
27-Jul	82	91	80	93	88	90	89	87	82	92	92	1.72	1.71544715	FALSE
28-Jul	79	91	88	93	91	84	91	89	86	90	94	4.72	6.43089430	FALSE
29-Jul	90	89	89	93	89	89	91	88	86	82	93	-6.28	0.14634146	FALSE
30-Jul	91	88	90	97	87	89	88	84	90	84	94	-7.28	0	FALSE
31-Jul	87	72	86	99	86	87	90	88	80	85	93	-3.28	0	FALSE
1-Aug	86	80	86	96	86	84	93	84	87	81	89	-2.28	0	FALSE
2-Aug	90	84	82	93	81	84	91	84	89	84	94	-6.28	0	FALSE
3-Aug	84	88	84	88	84	84	91	84	88	88	94	-0.28	0	FALSE
4-Aug	91	89	86	89	88	86	91	82	90	90	97	-7.28	0	FALSE

Figure 6.2.2 - Model Snapshot with C = 0

So I played around with the Dampener until I saw the S(1996) column become less fickle and shows lesser flopping..

The Dampener I ended up finalizing was 0.7 x standard deviation for the data for each year.

I then played around with the value of T:

Starting with T=2σ (or 2 times standard deviation) I notice that the change is detected on 30th September.

Variables															
Target (T)	2 standard deviations from norm														
Dampener (C)	0.7 standard deviations from norm														
AVERAGE:	83.72	81.67	84.26	83.36	84.03	81.55	83.59	81.48	81.67	83.94	83.30				
STDDEV:	8.55	9.32	6.41	9.72	9.52	8.22	9.43	7.02	7.73	6.59	8.71				
Target (T)	17.10	18.64	12.82	19.45	19.04	16.45	18.85	14.04	15.45	13.18	17.42				
Dampener (C)	5.98	6.52	4.49	6.81	6.66	5.76	6.60	4.91	5.41	4.61	6.10				
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?	
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-20.27	0	FALSE	
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-19.27	0	FALSE	
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-19.27	0	FALSE	
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-12.27	0	FALSE	
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-11.27	0	FALSE	
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-15.27	0	FALSE	
22-Sep	81	70	88	72	73	87	77	75	82	82	76	-3.27	0	FALSE	
23-Sep	84	80	84	75	81	88	82	81	82	77	81	-6.27	0	FALSE	
24-Sep	84	82	81	78	84	69	73	80	71	78	74	-6.27	0	FALSE	
25-Sep	87	66	82	81	82	66	69	82	67	77	67	-9.27	0	FALSE	
26-Sep	84	70	84	82	68	72	75	82	78	74	71	-6.27	0	FALSE	
27-Sep	79	64	87	78	71	75	75	82	79	78	71	-1.27	0	FALSE	
28-Sep	75	68	80	80	75	78	79	73	77	74	75	2.73	2.73160968	FALSE	
29-Sep	72	77	75	77	73	71	73	66	76	71	77	5.73	8.46321937	FALSE	
30-Sep	64	86	75	71	75	71	79	71	77	84	85	13.73	22.1948290	TRUE	
1-Oct	66	75	86	73	77	75	82	72	82	86	71	11.73	33.9264387	TRUE	
2-Oct	72	73	78	75	79	80	84	68	82	85	66	5.73	39.6580484	TRUE	
3-Oct	84	75	77	84	82	81	84	66	82	78	66	-6.27	33.3896581	TRUE	
4-Oct	70	78	82	71	81	80	82	77	85	65	70	7.73	41.1212678	TRUE	
5-Oct	66	81	82	73	82	79	87	78	84	71	73	11.73	52.8528775	TRUE	

Figure 6.2.3 - Model Snapshot with T=2 (C fixed at 0.7)

If I took T to 4σ or 2 times standard deviation) I notice that the change is detected on 30th September, the day of change didn't move that drastically (ie. it was Oct 2nd)

Variables															
Target (T)	4 standard deviations from norm														
Dampener (C)	0.7 standard deviations from norm														
AVERAGE:	83.72	81.67	84.26	83.36	84.03	81.55	83.59	81.48	81.67	83.94	83.30				
STDDEV:	8.55	9.32	6.41	9.72	9.52	8.22	9.43	7.02	7.73	6.59	8.71				
Target (T)	34.19	37.28	25.64	38.89	38.07	32.90	37.70	28.07	30.91	26.37	34.84				
Dampener (C)	5.98	6.52	4.49	6.81	6.66	5.76	6.60	4.91	5.41	4.61	6.10				
DAY	1996	1997	1998	1999	2000	2001	2002	2003	2013	2014	2015	u-x(1996)-C	S(1996)	Is S(1996) > T(1996) ?	
1-Jul	98	86	91	84	89	84	90	73	82	90	85	-20.27	0	FALSE	
2-Jul	97	90	88	82	91	87	90	81	85	93	87	-19.27	0	FALSE	
3-Jul	97	93	91	87	93	87	87	87	76	87	79	-19.27	0	FALSE	
4-Jul	90	91	91	88	95	84	89	86	77	84	85	-12.27	0	FALSE	
5-Jul	89	84	91	90	96	86	93	80	83	86	84	-11.27	0	FALSE	
6-Jul	93	84	89	91	96	87	93	84	83	87	84	-15.27	0	FALSE	
22-Sep	81	70	88	72	73	87	77	75	82	82	76	-3.27	0	FALSE	
23-Sep	84	80	84	75	81	88	82	81	82	77	81	-6.27	0	FALSE	
24-Sep	84	82	81	78	84	69	73	80	71	78	74	-6.27	0	FALSE	
25-Sep	87	66	82	81	82	66	69	82	67	77	67	-9.27	0	FALSE	
26-Sep	84	70	84	82	68	72	75	82	78	74	71	-6.27	0	FALSE	
27-Sep	79	64	87	78	71	75	75	82	79	78	71	-1.27	0	FALSE	
28-Sep	75	68	80	80	75	78	79	73	77	74	75	2.73	2.73160968	FALSE	
29-Sep	72	77	75	77	73	71	73	66	76	71	77	5.73	8.46321937	FALSE	
30-Sep	64	86	75	71	75	71	79	71	77	84	85	13.73	22.1948290	FALSE	
1-Oct	66	75	86	73	77	75	82	72	82	86	71	11.73	33.9264387	FALSE	
2-Oct	72	73	78	75	79	80	84	68	82	85	66	5.73	39.6580484	TRUE	
3-Oct	84	75	77	84	82	81	84	66	82	78	66	-6.27	33.3896581	FALSE	
4-Oct	70	78	82	71	81	80	82	77	85	65	70	7.73	41.1212678	TRUE	
5-Oct	66	81	82	73	82	79	87	78	84	71	73	11.73	52.8528775	TRUE	
6-Oct	64	82	73	71	73	70	86	75	84	78	76	13.73	66.5844872	TRUE	
7-Oct	60	82	82	73	66	68	80	73	74	82	81	17.73	84.3160968	TRUE	
8-Oct	78	82	69	73	55	79	71	73	72	86	82	0.27	84.0477085	TRUE	



Since 4 standard deviations is a considerable change of **DROP** in temperature, I decided to use this as my barometer (no pun intended)

When the final model was set up , I changed S>T cells color to green if S>T by using conditional formatting.

With  $T = 4 \times \text{STDEV}$ , and  $C = 0.7 \times \text{STDEV}$  (my original values) it is clear that unofficially temperature noticeably starts to cool down in the first half of October.

Figure 6.2.5 -  $S>T$

Interestingly enough, if I increase the dampener value , C to 1, the 2015 temperature change detection gets drastically modified

FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
FALSE	FALSE	FALSE
TRUE	TRUE	FALSE
TRUE	TRUE	FALSE
TRUE	TRUE	FALSE
TRUE	TRUE	FALSE
TRUE	TRUE	TRUE

*Figure 6.2.6 with bigger dampener ( $C=1$ )*

This tells us that temperature was drastically but steadily dropping in 2015 starting early October, but when the dampener has a high value, these changes are not detected.

Another interesting observation was that if we decrease the threshold  $T$  to  $2 \times \text{STDEV}$ , the change (to cooler season) is detected sooner, by about a week), but the year over year pattern doesn't change.

[illegible]



**Question 6.2.1 :Use a CUSUM approach to make a judgment of whether Atlanta’s summer climate has gotten warmer in that time (and if so, when).**

[illegible]

However to prove this via a CUSUM model... I first, counted how many days in before the change is detected by using COUNTIFS(range, FALSE)

=COUNTIFS(AS10:AS132, FALSE)										
A	B	C	V	W	AR	AS	AT	AU	AV	AW
Variables										
Target (T)	4	standard deviations from norm								
Dampener (C)	0.7	standard deviations from norm								
AVERAGE:	83.72	81.67								
STDDEV:	8.55	9.32								
Target (T)	34.19	37.28		S values	Is S>T ?					
Dampener (C)	5.98	6.52				1996	1997	1998	1999	2
DAY	1996	1997				Is S(1996) > T(1996) ?	Is S(1997) > T(1997) ?	Is S(1998) > T(1998) ?	Is (1999) > T(1999)	Is S(2000)>T(2000)
1-Jul	98	86				FALSE	FALSE	FALSE	FALSE	FALSE
2-Jul	97	90				FALSE	FALSE	FALSE	FALSE	FALSE
28-Oct	81	55				TRUE	TRUE	TRUE	TRUE	TRUE
29-Oct	82	64				TRUE	TRUE	TRUE	TRUE	TRUE
30-Oct	82	66				TRUE	TRUE	TRUE	TRUE	TRUE
31-Oct	81	60				TRUE	TRUE	TRUE	TRUE	TRUE
				Days in before winter(unoff)		94	107	100	103	

Figure 6.2.8 Summarizing yearly data (I'm hiding many of the rows to fit the snapshot in cleanly)

The above highlighted cell shows that it was 94 days after 7/1 when the change was detected (that is the temperature got notably cooler)

I transposed the table to make it line up against years , and then just like I did for each year, I calculated, the Average, Std Dev, T, and C for each of these years as one data set (I kept  $T=4 \times \text{STDEV}$  and  $C=0.7 \times \text{STDEV}$  like before)

I was then able to evaluate  $S(t)$  for each of the value, and compared that with T .

Year	Days after 7/1 before it gets cold	S(t)	Is S(t) > T ?
1996	94	0	FALSE
1997	107	4.771410584	FALSE
1998	100	0	FALSE
1999	103	0	FALSE
2000	99	0	FALSE
2001	103	0	FALSE



2002	106	0	FALSE
2003	93	0	FALSE
2004	104	5.77141 0584	FALSE
2005	114	0.54282 11678	FALSE
2006	107	0	FALSE
2007	112	0	FALSE
2008	111	0	FALSE
2009	107	0	FALSE
2010	94	0	FALSE
2011	102	4.77141 0584	FALSE
2012	100	1.54282 1168	FALSE
2013	110	0.31423 17517	FALSE
2014	106	0	FALSE
2015	93	0	FALSE
<b>AVG:</b>	<b>103.25</b>		
<b>STDEV</b>	<b>6.39798488</b>		
<b>Target (T)</b>	<b>25.59193952</b>		
<b>Dampener (C)</b>	<b>4.478589416</b>		

*Table 6.2.9 CUSUM model against changes to winter start Year over year.*

**Clearly, the temperature has \_not\_ gotten noticeably cooler over the years.**