hw3-6.1-2

Mark Pearl

01/29/2019

6.1 At work we have a number of clients in the oil & gas industry who need to have sophistication for the area of preventative maintenance with sensor data. Near-real time streaming data feeds into an enterprise data platform to conduct time-series analytics on the health of various machines across different oil sites. CUSUM is an applicable approach for solving this problem as the health of the machine needs to stay below a set threshold set from maintenance staff.

We want our approach to be conservative enough so we don't measure false alarms but tolerant enough to prevent major breakdowns. This would allude to use choosing a moderate level for our threshold.

https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/CUSUM_Charts.pdf

Based on the following resource we can see that suitable values for T is 5 and the value for c is usually set to 0.5. If downtimes were extremely critical then we would likely choose a lower threshold.

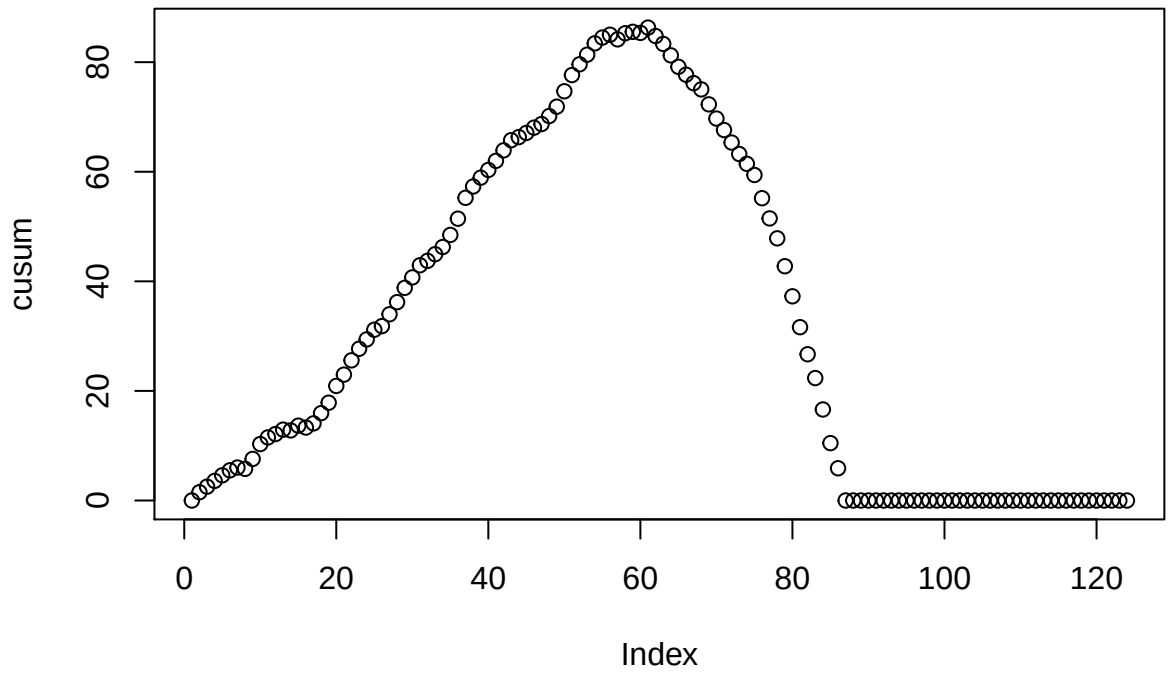
##6.2.1 Cusum Calculation: In this section we'll be calculating the cusum by calculating the mean for each day across all years from 1996 to 2015. We'll then iterate across each year and subtract the x of t and c from the mean as we're looking to detect change for a decrease in temperature.

```
temps_data <- read.table('C:/Users/mjpearl/Desktop/omsa/ISYE-6501-OAN/hw3/data/temps.txt',header = TRUE
head(temps_data)
```

```
##     DAY X1996 X1997 X1998 X1999 X2000 X2001 X2002 X2003 X2004 X2005 X2006
## 1 1-Jul    98    86    91    84    89    84    90    73    82    91    93
## 2 2-Jul    97    90    88    82    91    87    90    81    81    89    93
## 3 3-Jul    97    93    91    87    93    87    87    87    86    86    93
## 4 4-Jul    90    91    91    88    95    84    89    86    88    86    91
## 5 5-Jul    89    84    91    90    96    86    93    80    90    89    90
## 6 6-Jul    93    84    89    91    96    87    93    84    90    82    81
##   X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
## 1    95    85    95    87    92   105    82    90    85
## 2    85    87    90    84    94    93    85    93    87
## 3    82    91    89    83    95    99    76    87    79
## 4    86    90    91    85    92    98    77    84    85
## 5    88    88    80    88    90   100    83    86    84
## 6    87    82    87    89    90    98    83    87    84
```

```
#Calculate the average temperate for each across day across all years from 1995 to 2015
x_of_t <- rowMeans(temps_data[c(2:length(temps_data))], dims=1, na.rm=T)

#Compute the mean for all values of x of t to get mu
mu <- mean(x_of_t)

#Set our value of C to help moderate our values to ensure we don't go over the threshold
C <- 4

#Now to compute S of t to determine if we're experiencing a change below our threshold we use the formu
s_of_t <- x_of_t - mu - C
```

```r
#Craete an empty vector with an issue value of 0 but will be appended the value of cusum in a loop for
precusum <- 0 * s_of_t
cusum <- append(precusum, 0)

#Loop through all observations to update the cusum value and check to #see if it's negative. If it's ne
for (i in 1:length(s_of_t))
     {
  cusum_value <- cusum[i] + s_of_t[i]
  ifelse(cusum_value > 0, cusum[i+1] <- cusum_value, cusum[i+1] <- 0)
}
plot(cusum)
```




```r
which(cusum > 85)
```

```
## [1] 56 58 59 60 61
```

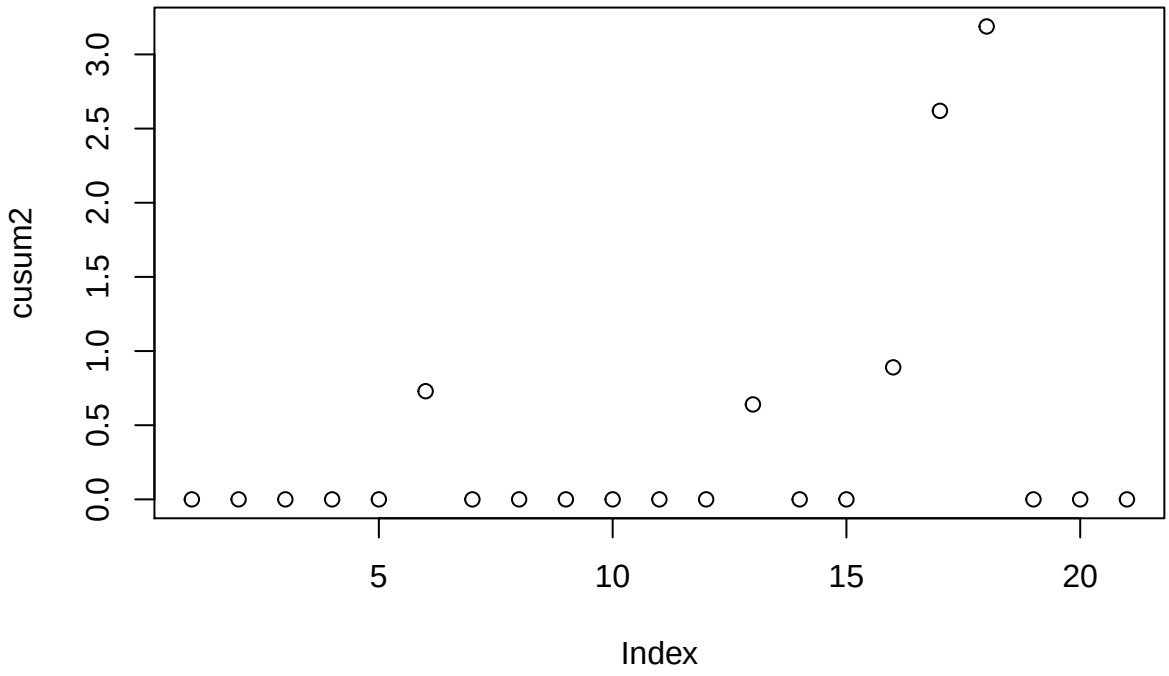
```r
temps_data[56, 1]
```

```
## [1] "25-Aug"
```

From our results we can determine that there is a temperate change starting from the 56th obversation/row in the dataset or August 25th onwards.

##6.2.2 Cusum calculation: Use a CUSUM approach to make a judgment of whether Atlanta's summer climate has gotten warmer in that time (and if so, when)

```r
#Calculate the average temperate for each year up until we see a drop in the change of weather measured
temps_subset <- temps_data[1:56,]
x_of_t2 <- colMeans(temps_subset[c(2:length(temps_subset))], dims=1, na.rm=T)

#Compute the mean for all values of x of t to get mu
mu2 <- mean(x_of_t2)

#Set our value of C to help moderate our values to ensure we don't go over the threshold
C <- 2

#Now to compute S of t to determine if we're experiencing a change below our threshold we use the formu
s_of_t2 <- x_of_t2 - mu2 - C

#Craete an empty vector with an issue value of 0 but will be appended the value of cusum in a loop for
precusum2 <- 0 * s_of_t2
cusum2 <- append(precusum2, 0)

#Loop through all observations to update the cusum value and check to #see if it's negative. If it's ne
for (i in 1:length(s_of_t2))
     {
  cusum_value2 <- cusum2[i] + s_of_t2[i]
  ifelse(cusum_value2 > 0, cusum2[i+1] <- cusum_value2, cusum2[i+1] <- 0)
}
plot(cusum2)
```



From this result we can conclude that we are detecting changes above the threshold on several occasions

throughout this time interval. More so in the later years. However, it doesn't seem to follow a consistent trend and the results are inconclusive of whether the climate has got warmer.