

Clarifying the CUSUM function from the qcc package

So I believe I've finally gotten a true understanding of the cusum function, there seems to be a lot of confusion on what the arguments `se.shift` and `decision.interval` are used for. So here goes my best (at the moment) guess/comprehension. First, this function does NOT give you a raw cumulative sum, but instead gives it as a normalized value (that is, divided by the standard deviation). If your data set is in degrees F for units, it does not output the cusum in degrees F, it does so in standard deviations.

Description

Create an object of class 'cusum.qcc' to compute a Cusum chart for statistical quality control.

Usage

```
cusum(data, sizes, center, std.dev, head.start = 0,
decision.interval = 5, se.shift = 1, data.name, labels,
newdata, newsizes, newlabels, plot = TRUE, ...)
```

The pertinent arguments to enter are data, center, std.dev, decision.interval, and se.shift.

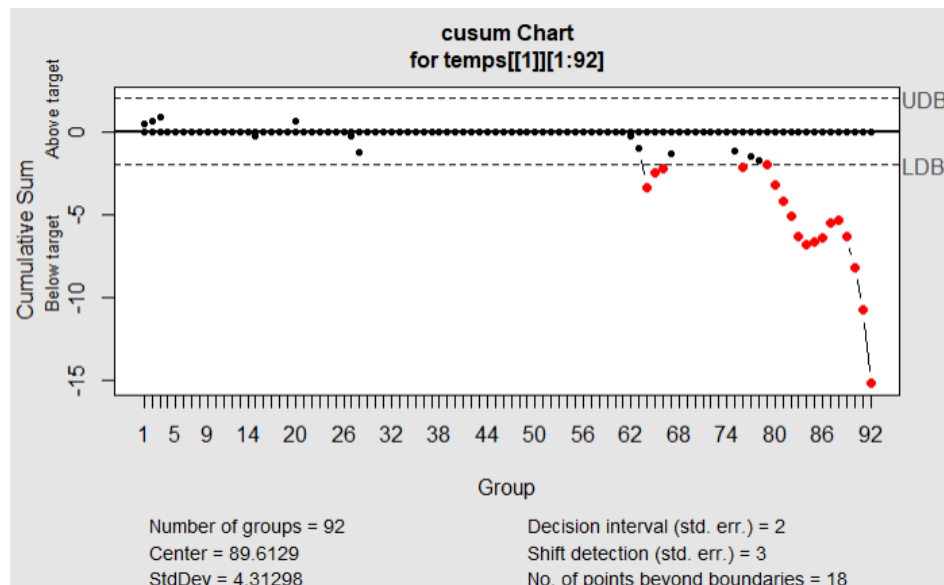
I've done some digging through the code and worked out the math and the St equation is:

$$St = \max \left\{ 0, St_0 + \frac{(Value - \mu - \frac{SD \cdot Shift}{2})}{SD} \right\}$$

Where value is the observed value for that point in time, mu is obviously your mu (or center for the function), and SD is the standard deviation you choose (either from the data set or another value you want).

So if you want $C = 1.5 * SD$ and a threshold $= 2 * SD$ s (again this would be a cumulative sum of 2 times the SD you set), the code would be:

```
temp <- cusum(temps[[1]][1:92],
              center = avg_summer[1],
              std.dev = sd_summer[1],
              decision.interval = 2,
              se.shift = 3,
              plot = TRUE)
```



If you'd want a specific total temperature below or above for your C and threshold, you'd have to do them in multiples of your SD.

So say I wanted a $C = 4F$ and $T = 20F$:

I could set `std.dev = 4`, `se.shift = 2`, and `T = 5`

$$St = \max \left\{ 0, St_0 + \frac{(Value - \mu - \frac{4 \cdot 2}{2})}{4} \right\}$$

Setting `se.shift = 2` cancels out the division that occurs.

#edited: added max(0, formula). Thanks, Sarnath.
#pin

hw2 hw3 week2 week3

Updated 1 day ago by Matthew Nguyen

followup discussions for lingering questions and comments

☒ Resolved ☐ Unresolved



Sarnath Kannan 1 day ago

Thanks for the note. It helps understand the API... Since the API is chosen to represent C and T as multiples of SD, the equation too appears so. I think that makes sense. $(X - \mu)/SD$ measures number of Standard Deviations "X" is away.

Also to remember that the original formula from the lecture uses $\max(0, St)$ so if this values goes down to -ve, cumulation is held at 0 until the value emerges out of its negative trot...

Again, Thanks for caring to post this. This is a very useful note. I used my own method and made mistakes in my Homework (I did not use this API).... So, I had a feeling that I missed on this API.... Your note helped. Thanks!



Matthew Nguyen 1 day ago Thanks for the reminder re: max. I've updated it.

I made my own function for the last hw, but decided to give this a try. Also easier to incorporate code and do less of the heavy lifting that someone's already done.



Sarnath Kannan 1 day ago

Yeah, I too did... Just that my function was wrong.. I missed out the max :)



Omer Ansari 13 hours ago Thanks Matthew for digging into the code!

3 questions for you:

1. what really is `se.shift`? Whats really blowing my mind is that I read `se.shift` from the man page looks like "**T**" itself

`se.shift` The amount of shift to detect in the process, measured in standard errors of the summary statistics.

from your writing, it has more to do with "C" ? Can you explain your example where you want to use $C = 1.5 * SD$, and you specify an `se.shift` of 3?

later in a 2nd example you suggest $C = 4F$, and set `se.shift` = 2..

2. where do you use T ? in your 2nd example you want to use $T = 20F$, for this you suggest using $T=F$ (to quote: **I could set std.dev = 4, se.shift = 2, and T = 5**) but eventually i dont see a 5 anywhere in your equation, to quote:

$St = \max(0, St0 + (Value - \mu - (4 * 22)4)$

3. how do you write those fancy math equations with the symbols? That's so cool! When I copy pasted your equation it turned out flat as you can see above..



Matthew Nguyen 5 minutes ago Omer, I'm not writing the below with 100% certainty, just my interpretation.

1. `se.shift` essentially calculates the number of standard deviations you want to use for C, but splits this across the mu. So if your SD is 3 and mu is = 0, then having `se.shift` = 1, means you get an interval of 0 +/- 1.5. Setting `se.shift` = 2, would give an interval of 0 +/- 3. So essentially, setting `se.shift` = 2 gives you 1 SD above or below, and 3 would mean 1.5 above or below.

2. T is decision.interval, but again, in standard deviations. So if my standard deviation is 4 and I want to a $T = 20 F$, I need to divide T by SD to get the number of SDs I want me T to be above or below mu --> $20 / 4 = 5$.

3. When you write your response, use the LaTeX editor (the 'fx' button in the toolbar of your message).



Omer Ansari Just now Got it. This was quite helpful Matthew!