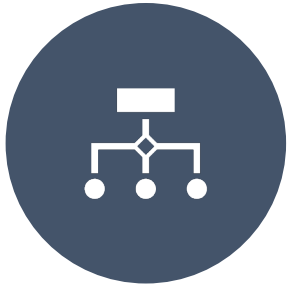
A Study on Variation of Number of Asthmatic Patients Admitted using Control Charts

Presentation by
Arvind Ramkumar

Data Description

A healthcare dataset is used to Explain the application of Control Chart in Healthcare.

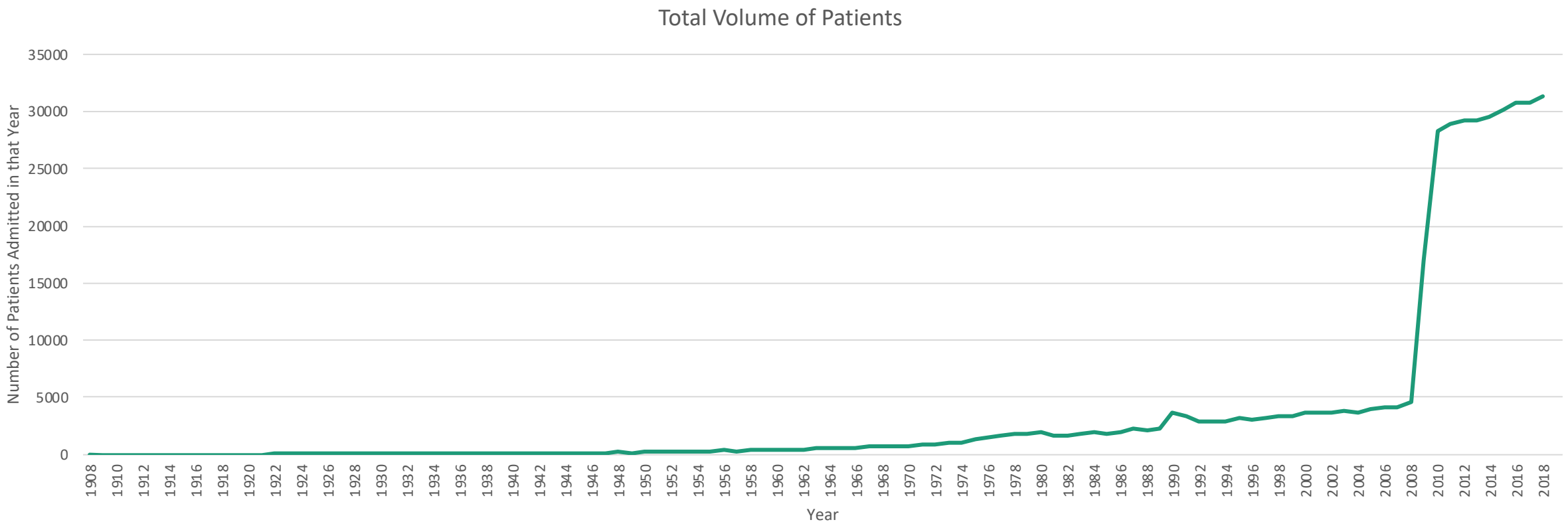


The dataset was provided as a part of the recruiting process to test my programming skills for this same position.

Only 10% of the data were used from the given for the first round, so the plan was to manipulate the data to describe how process Variation is measured and analyzed in Healthcare using SPC

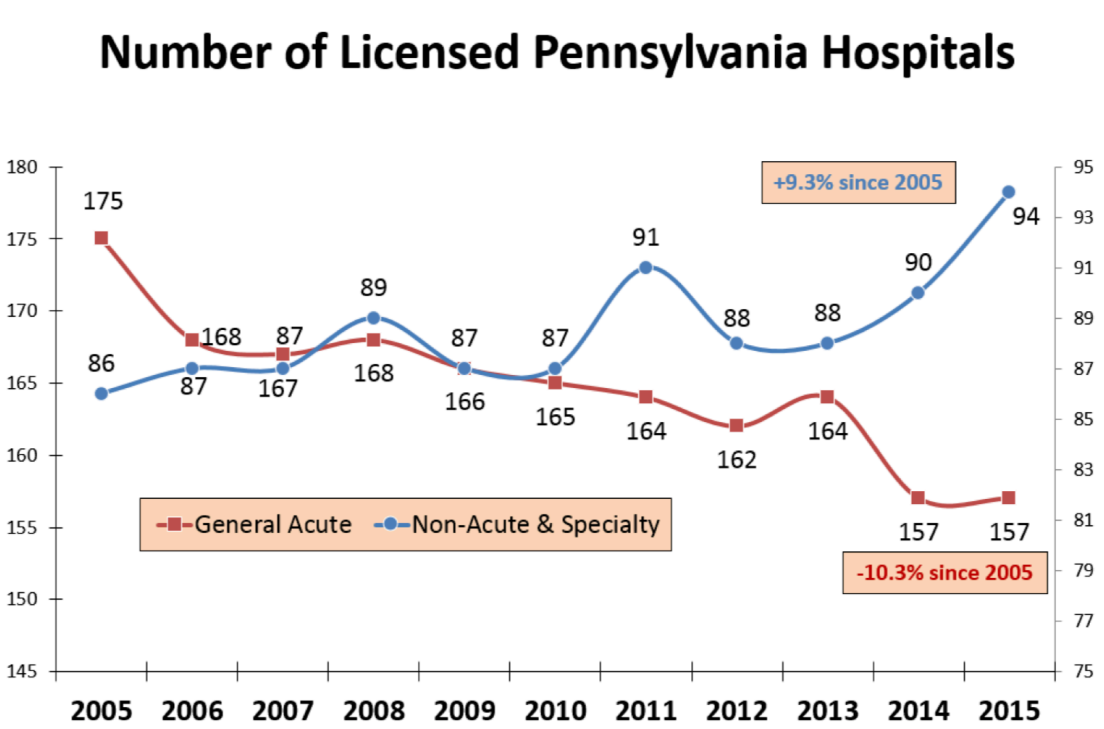
The data is gathered from a hospital in Pennsylvania

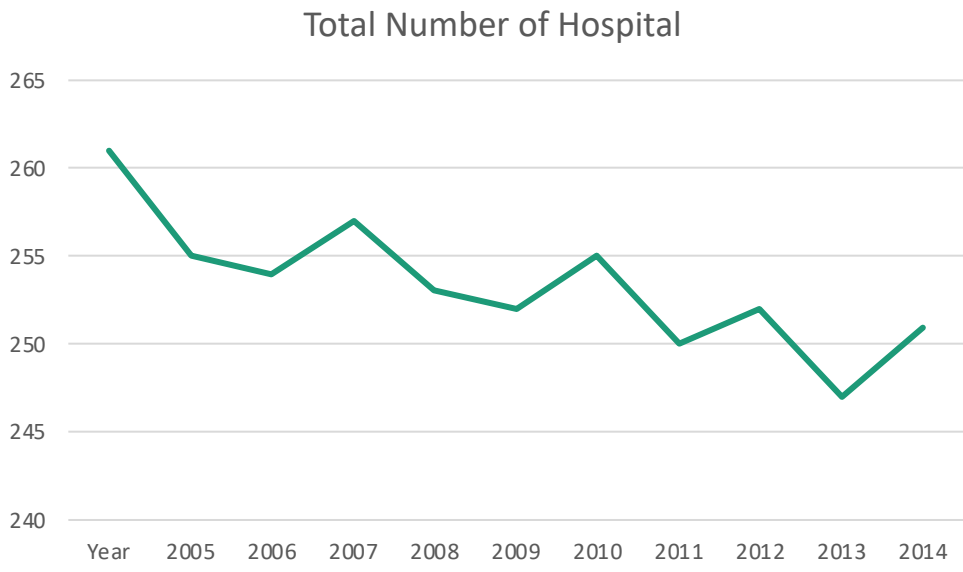
Total Volume of Patients coming to the Hospital



Number of Hospitals in Philadelphia

Source : Pennsylvania Department of Health





Possible Reasons for Sudden Rise in Patient Count at 2008

- Change in Data Collection Process and more diseases could have been accounted for
- The size of the Hospital could have been increases by size, volume, beds and more specialized treatments
- Number of hospitals in Pennsylvania could have been decreased, so more patients are inclined towards getting treatment in that specific hospital.
- Recession could have played a major role. And Children Hospitals were never hit in the recession, and the recession could have been used to capitalize and increase the size of the Organization.
- Population in Philadelphia could have been increased by a large volume during 2008 (But highly unlikely)

Objective

- To understand whether there was an increase in the flow of patients getting admitted in the hospital on account of Asthma (Between 2010 – 2018) and if so, understanding the assignable and special variation cause for this effect
- Why 2010 – 2018?
  - Because of a possible reason of recession (which we have established earlier), it is quite unfair to check for the process variation before and after.
  - There is a very obvious increase (in 2008) in the number of patients flow.
  - So we are concerned about whether there is any variation between 2010 and 2018.

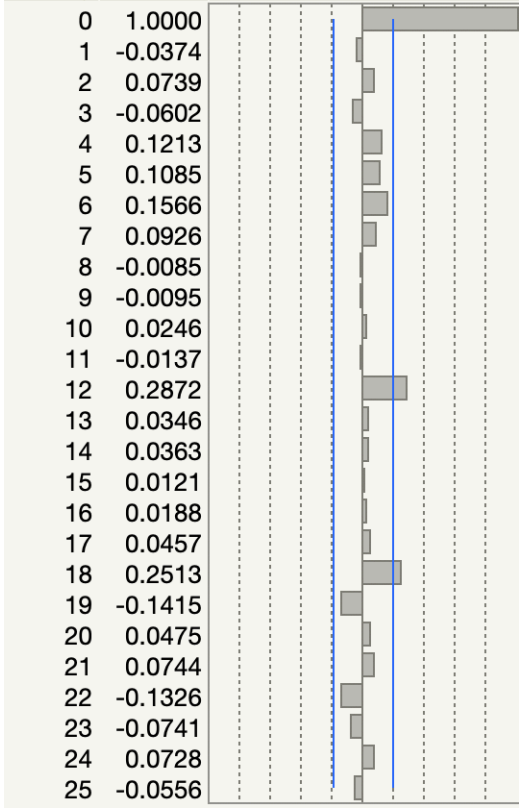
Type of Control Chart to be used?

- We use c – chart, and the reasons are specified below.
- The data is discrete and the underlying population is large (It is from a large pool of population). Also, the sample size is assumed to be constant (Here, the population in Philadelphia).
- Assumptions in c – chart
  - Events are relatively low frequency (In comparison with the underlying population)
  - Size of the underlying population is unknown
  - Independence
  - Normality

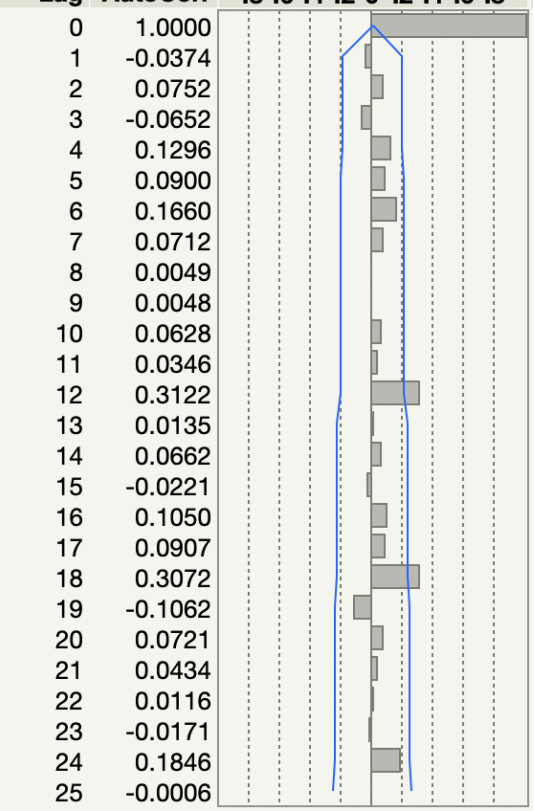
Check for Independence

All the lags of the residual is lying within the confidence band. So, the Independence Assumption is Valid

Lag	Partial
0	1.0000
1	-0.0374
2	0.0739
3	-0.0602
4	0.1213
5	0.1085
6	0.1566
7	0.0926
8	-0.0085
9	-0.0095
10	0.0246
11	-0.0137
12	0.2872
13	0.0346
14	0.0363
15	0.0121
16	0.0188
17	0.0457
18	0.2513
19	-0.1415
20	0.0475
21	0.0744
22	-0.1326
23	-0.0741
24	0.0728
25	-0.0556

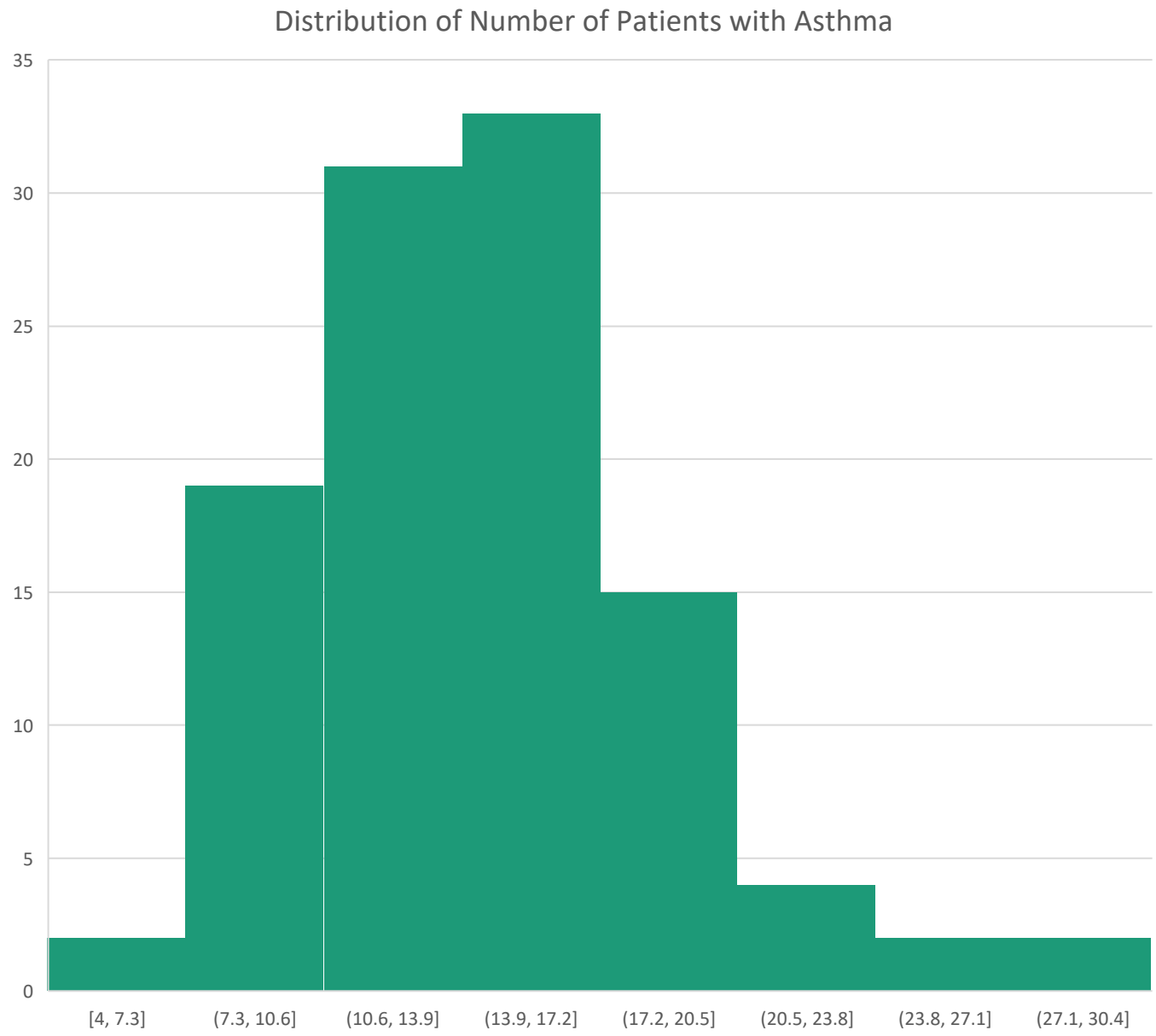


Lag	AutoCorr
0	1.0000
1	-0.0374
2	0.0752
3	-0.0652
4	0.1296
5	0.0900
6	0.1660
7	0.0712
8	0.0049
9	0.0048
10	0.0628
11	0.0346
12	0.3122
13	0.0135
14	0.0662
15	-0.0221
16	0.1050
17	0.0907
18	0.3072
19	-0.1062
20	0.0721
21	0.0434
22	0.0116
23	-0.0171
24	0.1846
25	-0.0006

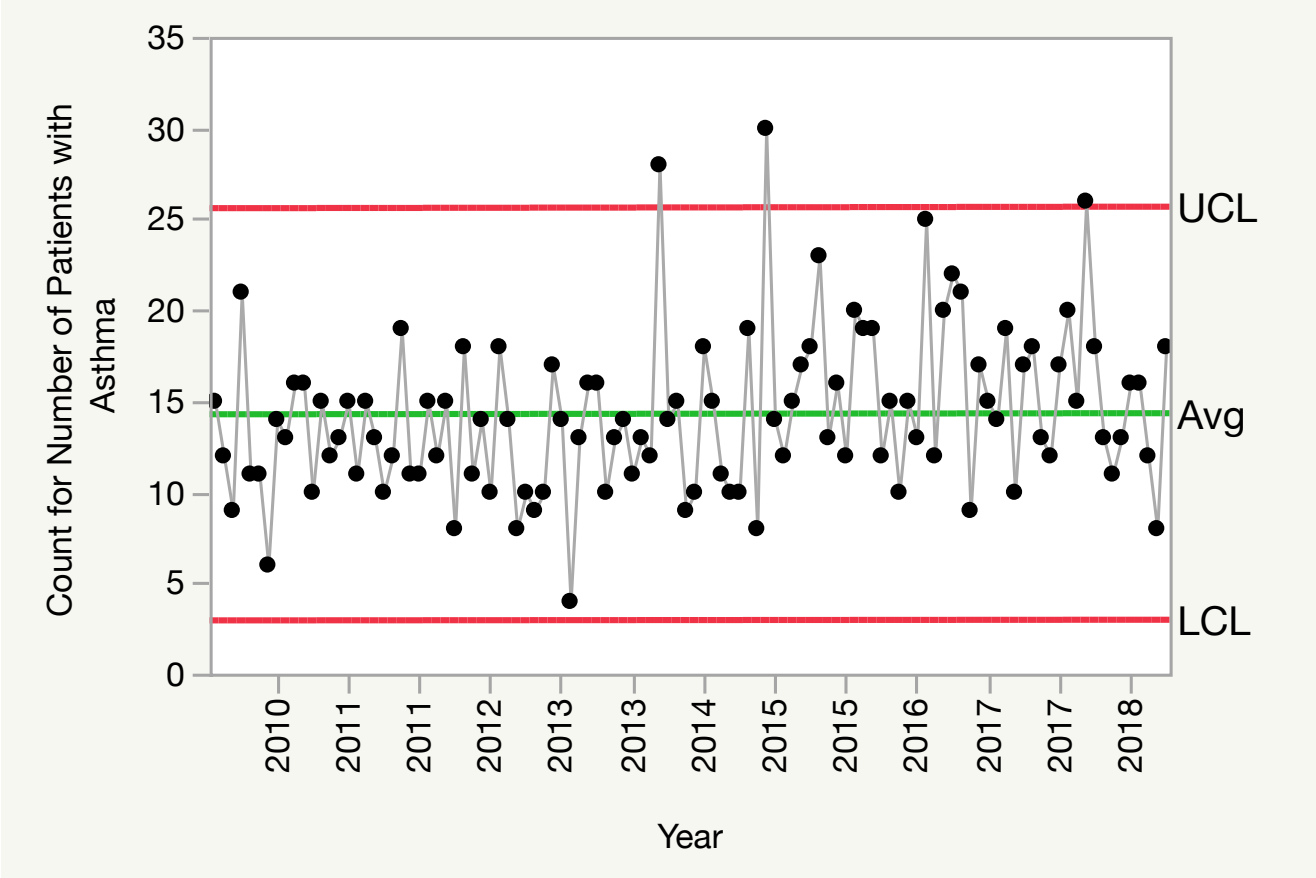


Check for Normality

The bell shaped curved will help
Us understand that the Normality
Assumption is valid or
at least near Normality



C – Chart

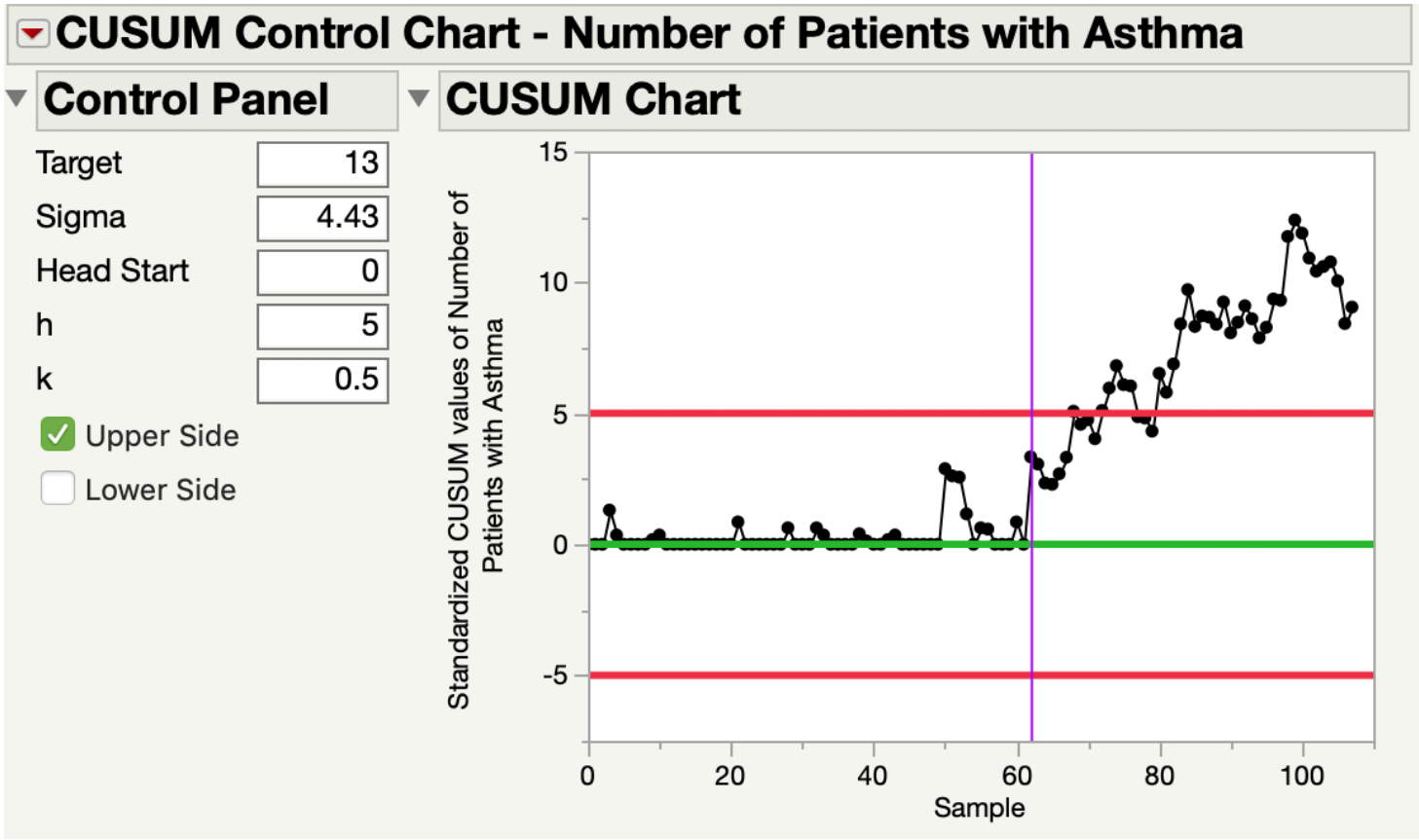


- Although there is only a couple of months that goes out control, there seems to be a very small mean change between 2010 – 2013 and 2014 – 2018.
- In general, it is impossible to determine a very small mean shift of probably 1σ with the Shewhart Control Chart
- During these times, to determine the very small mean shift, we can use CUCUM Chart

Parameter Selection in CUSUM Chart

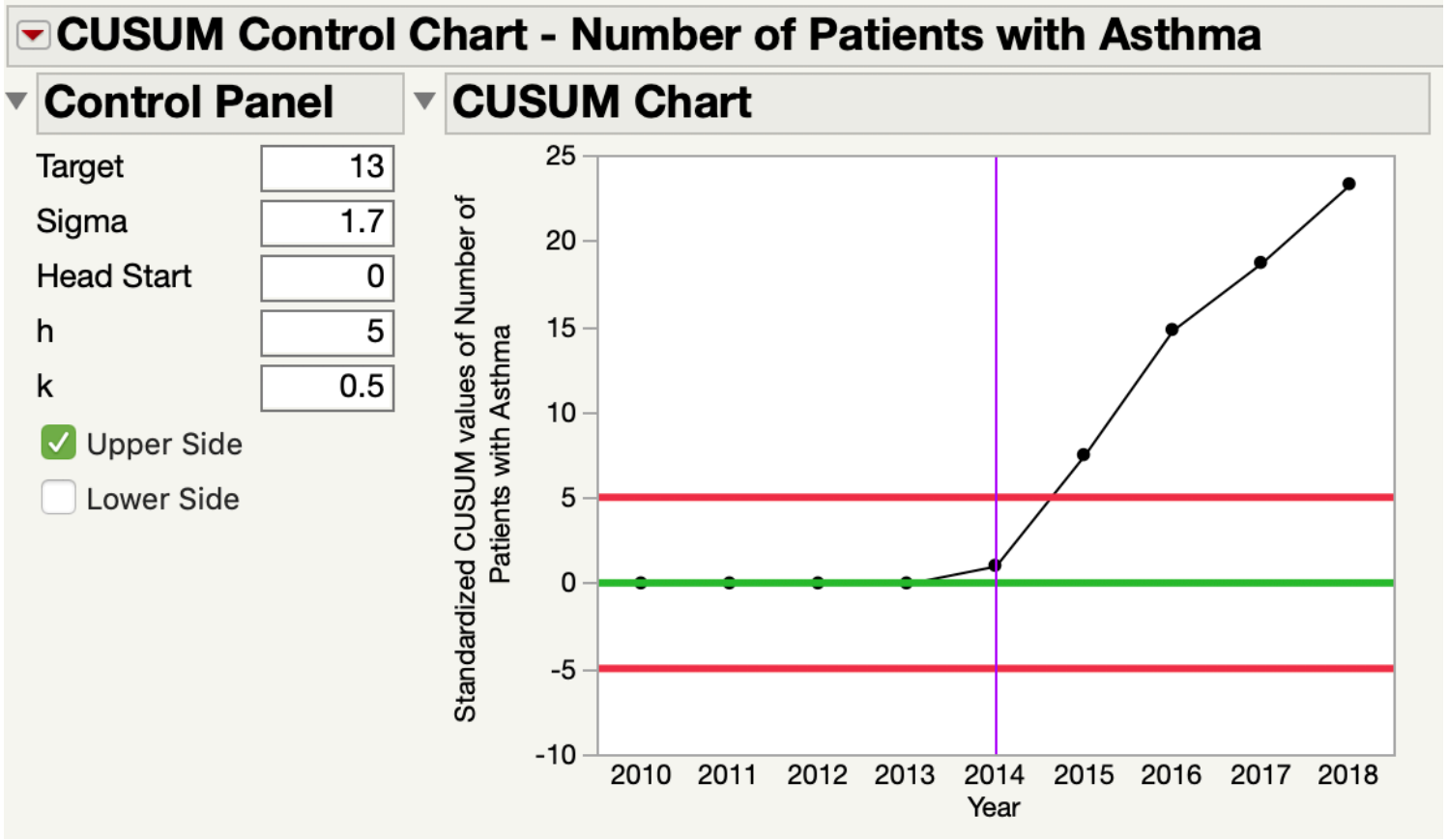
- Target – The Mean that we intend to obtain.
- σ – Usually the Population Standard Deviation. But here, since we don't have the population Standard Deviation, we use the estimate value (Sample Standard Deviation)
- k – Reference Value, specified in σ units. K is often set to half the change that needs to be detected. Here, since we are concerned about 1 σ change, we use k = 0.5
- H – Decision Limits, specified in σ units. Usually the value 5 is used. Of the curve goes beyond this Limit, then there is a definite mean shift in process.

CUSUM Chart – Individual Observation (2010 – 2018)



- From the above graph, we see that around 60 data points, the process starts to go out of control. So, after 2014 the process goes out of control. We now look at another CUSUM Control chart, the data points averaged over Year to help us clearer picture where the process goes out of control. This will help us determine the year at which the average number of patients has increases.

CUSUM Chart – Averaged Over Year (2010 – 2018)



- The Number of Patients is averaged over Year and a control chart is built for that average with the target Average to be 13 and we are looking for 1σ Change.
- Here, it makes us more clear that, beyond 2014, there is a definite Mean Shift.

Factors causing the mean shift

- Given, we have established that there is a mean shift in the number of patients, it is highly necessary to understand why there was this mean shift.
- If we are able to understand this mean shift, it will be possible for us to estimate what might happen in the future based on the trend that it follows.
- Two types of Causes
  - Chance Cause
  - Assignable Cause

Chance Cause and Assignable Cause

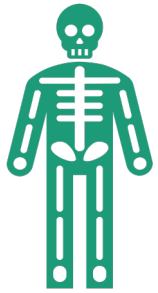
Chance Cause

Inherent part of a process.

Natural Patterns, Historical and Quantifiable

Here, Amount of Ozone, CO, SO2 and NO2 in atmosphere in Pennsylvania could be a chance cause for mean shift in Number of Patients.

These are parameters that can't be controlled, but causes an inevitable effect



Assignable Cause

Generally which wasn't observed before

Identifiable special cause of variation

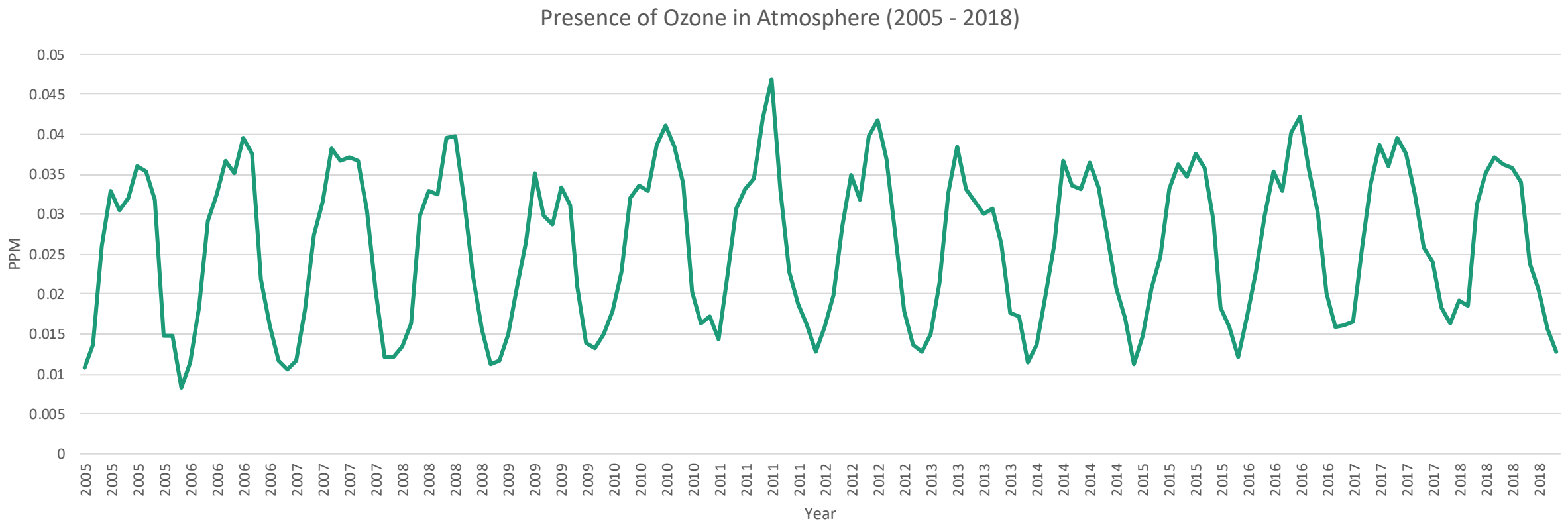
Here, during that period (2008) more specialized units to treat Asthma could have been established.

These parameters, might or might not be controllable.

Chance Cause

- Amount of Ozone in Atmosphere
- Amount of SO2 in Atmosphere
- Amount of CO in Atmosphere

Amount of Ozone in Atmosphere

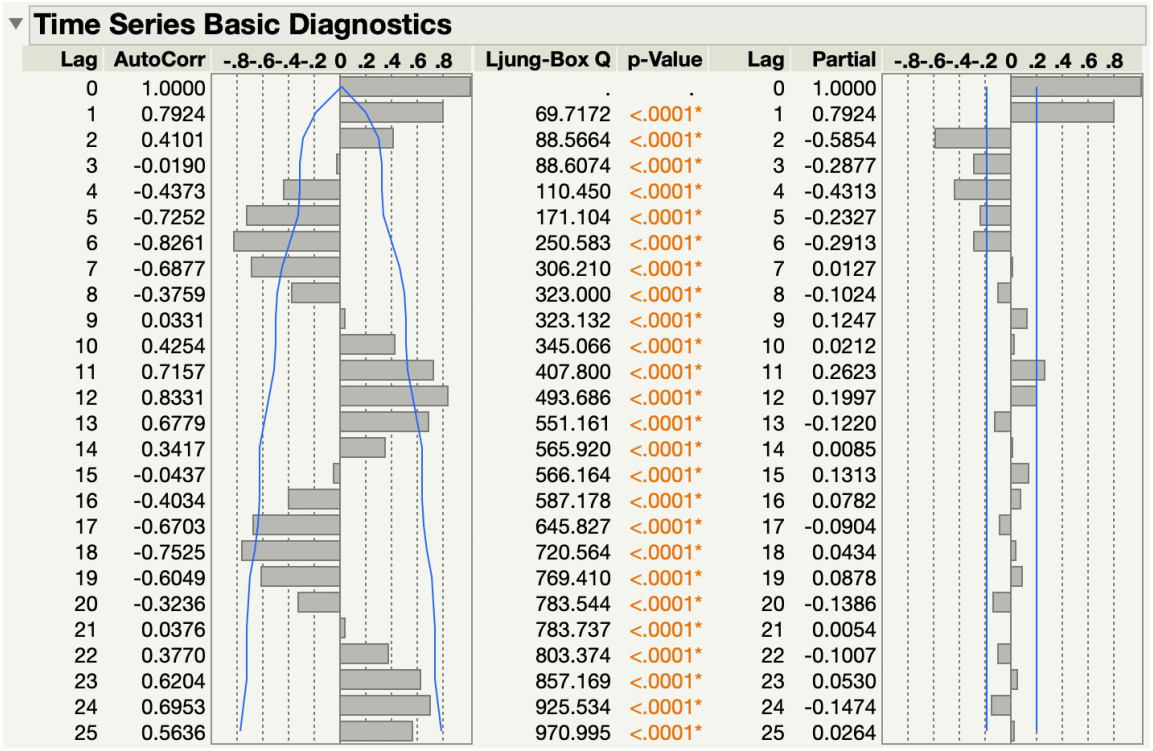


Type of Control Chart to be used?

- We use X – MR (I – MR) – chart, and the reasons are specified below.
- It is a continuous data (Parts Per Million is a continuous Variable) and the Sample size is 1. So we use this chart for determining the variation.
- Assumptions in X – MR Chart
  - One Observation per time period
  - Independence
  - Normality

Check for Independence

All the lags of the residual is lying above the confidence band. So, the Independence Assumption is Void



Time Series Basic Diagnostics

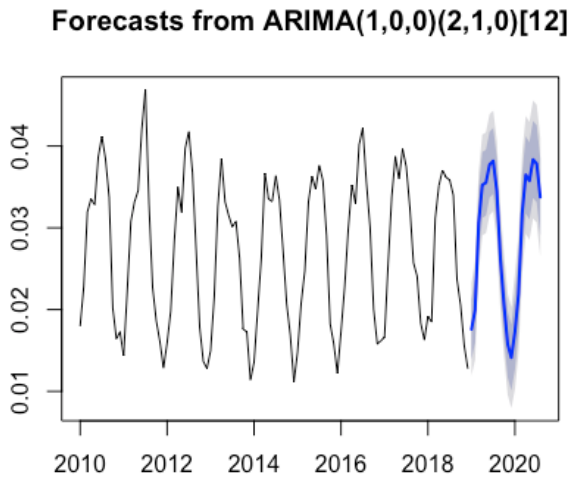
Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.7924		69.7172	<.0001*	1	0.7924	
2	0.4101		88.5664	<.0001*	2	-0.5854	
3	-0.0190		88.6074	<.0001*	3	-0.2877	
4	-0.4373		110.450	<.0001*	4	-0.4313	
5	-0.7252		171.104	<.0001*	5	-0.2327	
6	-0.8261		250.583	<.0001*	6	-0.2913	
7	-0.6877		306.210	<.0001*	7	0.0127	
8	-0.3759		323.000	<.0001*	8	-0.1024	
9	0.0331		323.132	<.0001*	9	0.1247	
10	0.4254		345.066	<.0001*	10	0.0212	
11	0.7157		407.800	<.0001*	11	0.2623	
12	0.8331		493.686	<.0001*	12	0.1997	
13	0.6779		551.161	<.0001*	13	-0.1220	
14	0.3417		565.920	<.0001*	14	0.0085	
15	-0.0437		566.164	<.0001*	15	0.1313	
16	-0.4034		587.178	<.0001*	16	0.0782	
17	-0.6703		645.827	<.0001*	17	-0.0904	
18	-0.7525		720.564	<.0001*	18	0.0434	
19	-0.6049		769.410	<.0001*	19	0.0878	
20	-0.3236		783.544	<.0001*	20	-0.1386	
21	0.0376		783.737	<.0001*	21	0.0054	
22	0.3770		803.374	<.0001*	22	-0.1007	
23	0.6204		857.169	<.0001*	23	0.0530	
24	0.6953		925.534	<.0001*	24	-0.1474	
25	0.5636		970.995	<.0001*	25	0.0264	

- Since the value is highly correlated and the independence assumption is violated very badly, the results that we might obtain out of the X MR Chart would be unreliable
- A general approach is we build a time series model and we build a control chart for the residuals, because it carries the same information and it will not be correlated and be independent of the observation

```
 ARIMA(2,0,2)(1,1,1)[12] with drift         : Inf
 ARIMA(0,0,0)(0,1,0)[12] with drift         : -811.605
 ARIMA(1,0,0)(1,1,0)[12] with drift         : -829.2863
 ARIMA(0,0,1)(0,1,1)[12] with drift         : Inf
 ARIMA(0,0,0)(0,1,0)[12]                    : -813.1284
 ARIMA(1,0,0)(0,1,0)[12] with drift         : -823.7349
 ARIMA(1,0,0)(2,1,0)[12] with drift         : -836.5459
 ARIMA(1,0,0)(2,1,1)[12] with drift         : Inf
 ARIMA(0,0,0)(2,1,0)[12] with drift         : -821.7814
 ARIMA(2,0,0)(2,1,0)[12] with drift         : -835.057
 ARIMA(1,0,1)(2,1,0)[12] with drift         : -835.4431
 ARIMA(2,0,1)(2,1,0)[12] with drift         : Inf
 ARIMA(1,0,0)(2,1,0)[12]                    : -838.6391
 ARIMA(1,0,0)(1,1,0)[12]                    : -831.2166
 ARIMA(1,0,0)(2,1,1)[12]                    : Inf
 ARIMA(0,0,0)(2,1,0)[12]                    : -823.7213
 ARIMA(2,0,0)(2,1,0)[12]                    : -837.1897
 ARIMA(1,0,1)(2,1,0)[12]                    : -837.5693
 ARIMA(2,0,1)(2,1,0)[12]                    : Inf

 Best model: ARIMA(1,0,0)(2,1,0)[12]

Series: myts
ARIMA(1,0,0)(2,1,0)[12]

Coefficients:
         ar1     sar1     sar2
      0.4089  -0.3944  -0.3580
s.e.  0.0941   0.1102   0.1081

sigma^2 estimated as 8.486e-06:  log likelihood=423.54
AIC=-839.08   AICc=-838.64   BIC=-828.82
```

Time Series Model for Ozone

- We use the ARIMA Model (1, 0, 0) (2, 1, 0) [12].
- The forecast graph is given below



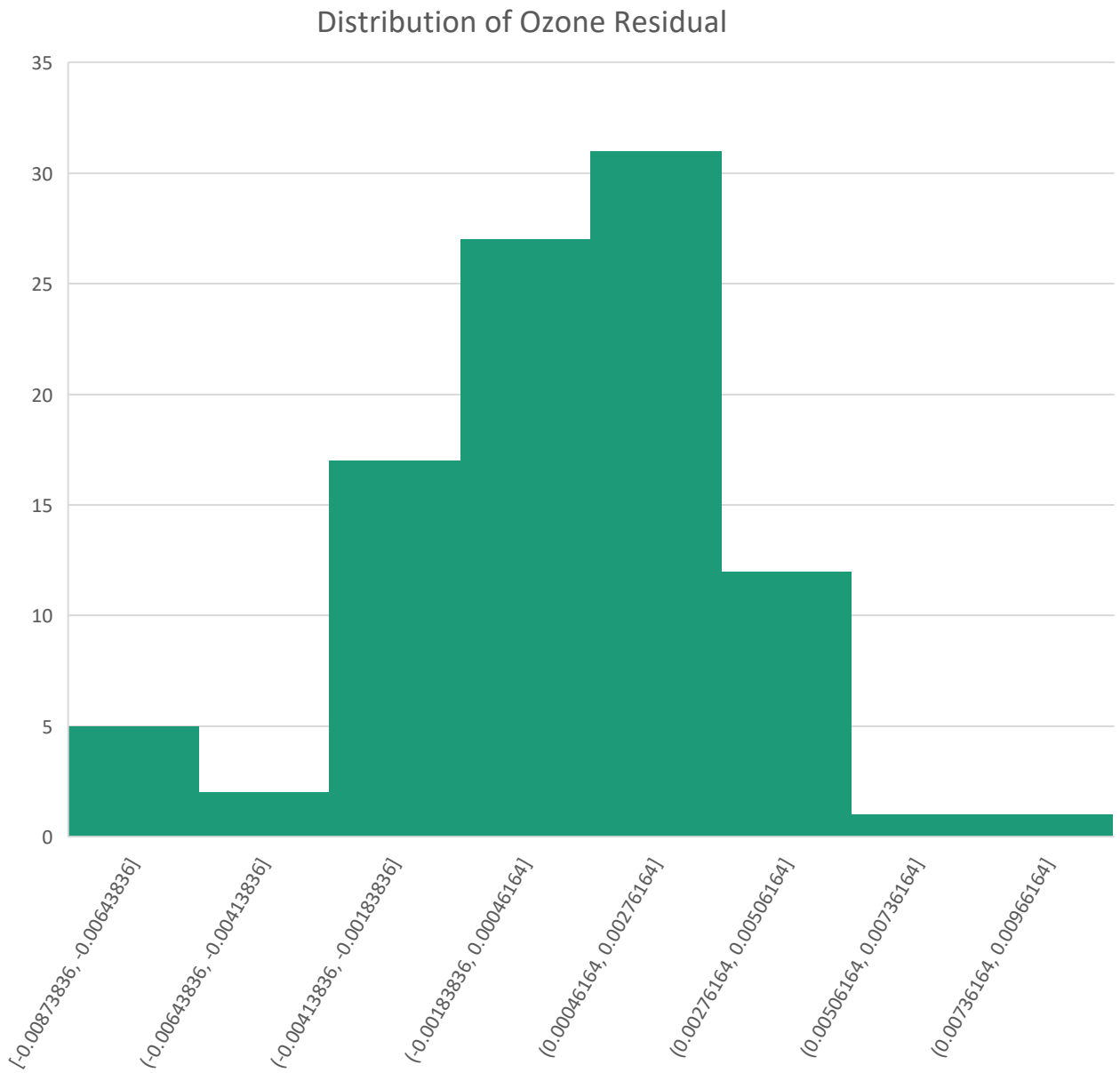
Independence in the Residuals

All the lags of the residual is lying within the confidence band. So, the Independence Assumption is Valid

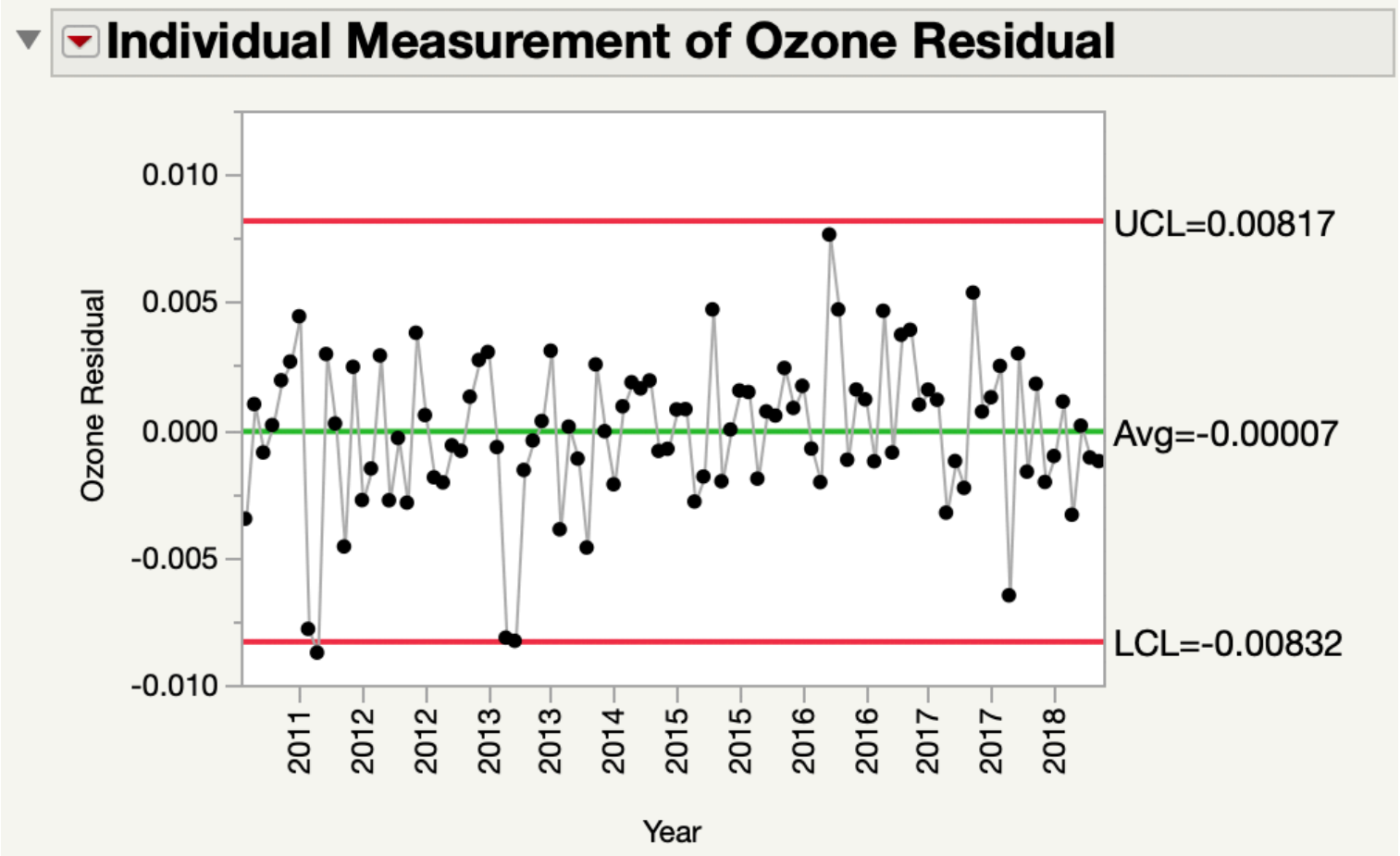
Lag	AutoCorr	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0389		0.1496	0.6989	1	0.0389	
2	-0.0959		1.0697	0.5858	2	-0.0975	
3	0.0393		1.2263	0.7467	3	0.0478	
4	-0.1535		3.6368	0.4574	4	-0.1690	
5	0.0572		3.9753	0.5530	5	0.0858	
6	0.0943		4.9041	0.5562	6	0.0516	
7	0.1333		6.7832	0.4518	7	0.1640	
8	0.1321		8.6495	0.3727	8	0.1056	
9	0.0653		9.1104	0.4271	9	0.1116	
10	0.0501		9.3846	0.4960	10	0.0812	
11	-0.0089		9.3933	0.5856	11	0.0349	
12	-0.0751		10.0248	0.6138	12	-0.0628	
13	0.1017		11.1977	0.5943	13	0.0968	
14	-0.0422		11.4023	0.6542	14	-0.1134	
15	0.0114		11.4173	0.7225	15	-0.0054	
16	0.0618		11.8669	0.7531	16	-0.0525	
17	0.0376		12.0354	0.7980	17	0.0527	
18	-0.0144		12.0605	0.8441	18	-0.0829	
19	-0.0428		12.2844	0.8731	19	-0.0229	
20	-0.0651		12.8092	0.8854	20	-0.1006	
21	0.1544		15.7984	0.7809	21	0.2118	
22	0.2207		21.9928	0.4603	22	0.2025	
23	-0.0297		22.1062	0.5138	23	0.0235	
24	-0.0861		23.0752	0.5154	24	-0.0978	
25	-0.0597		23.5473	0.5456	25	0.0073	

Normality of Ozone Residual

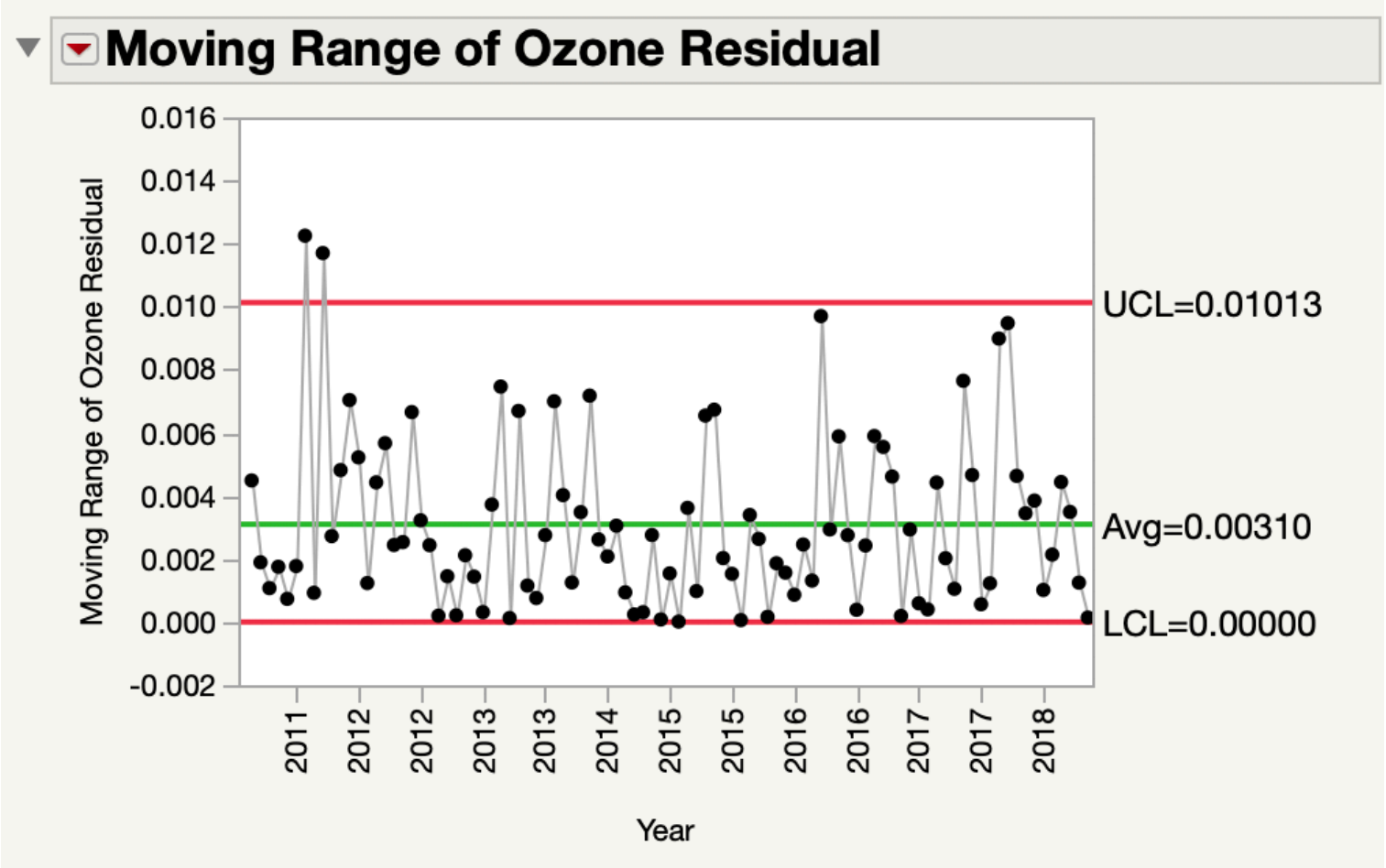
The bell shaped curved will help Us understand that the Normality Assumption is valid or at least near Normality



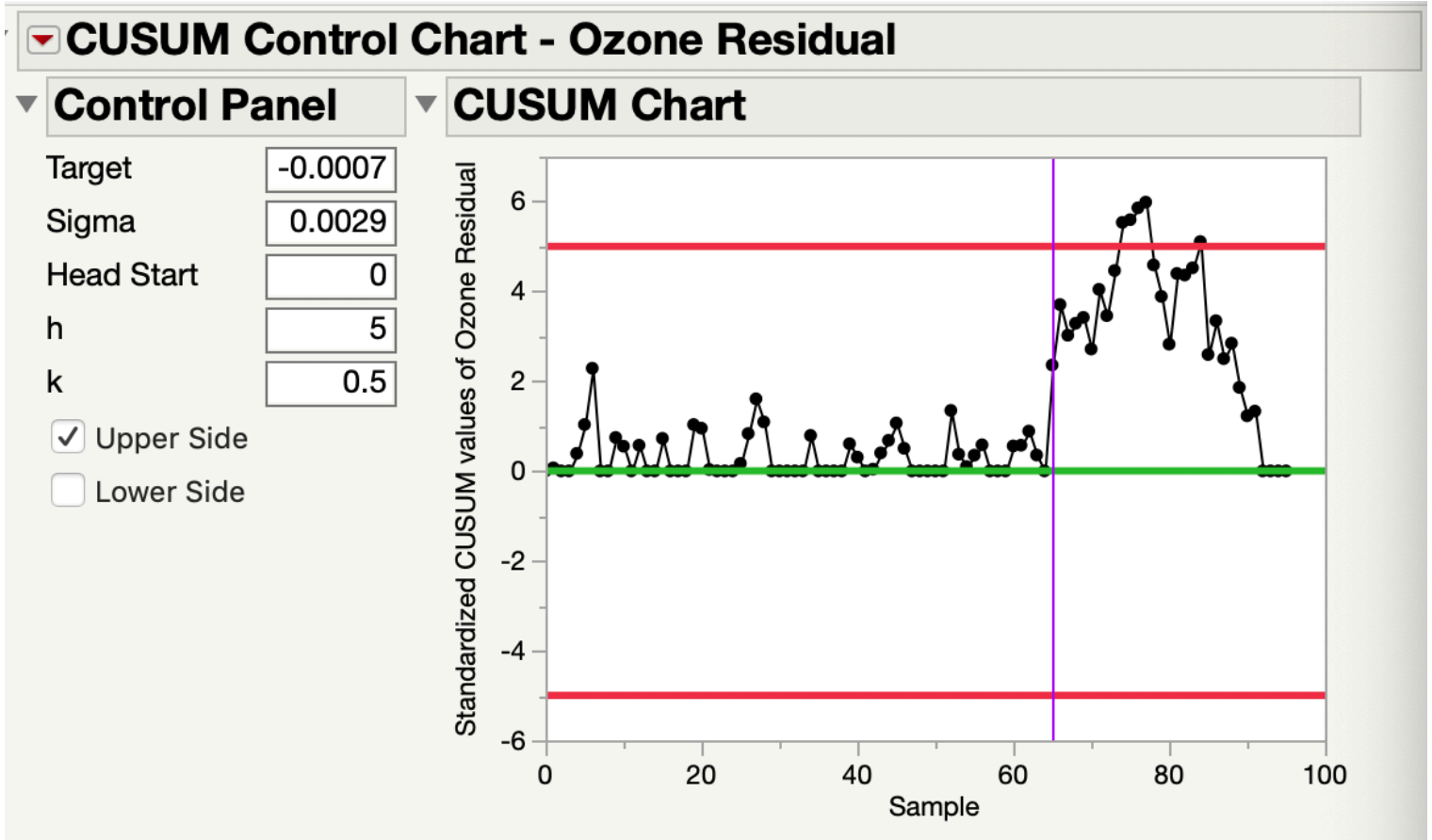
X/I Chart


Individual Measurement of Ozone Residual
Ozone Residual
0.010
0.005
0.000
-0.005
-0.010
UCL=0.00817
Avg=-0.00007
LCL=-0.00832
2011
2012
2012
2013
2013
2014
2015
2015
2016
2016
2017
2017
2018
Year

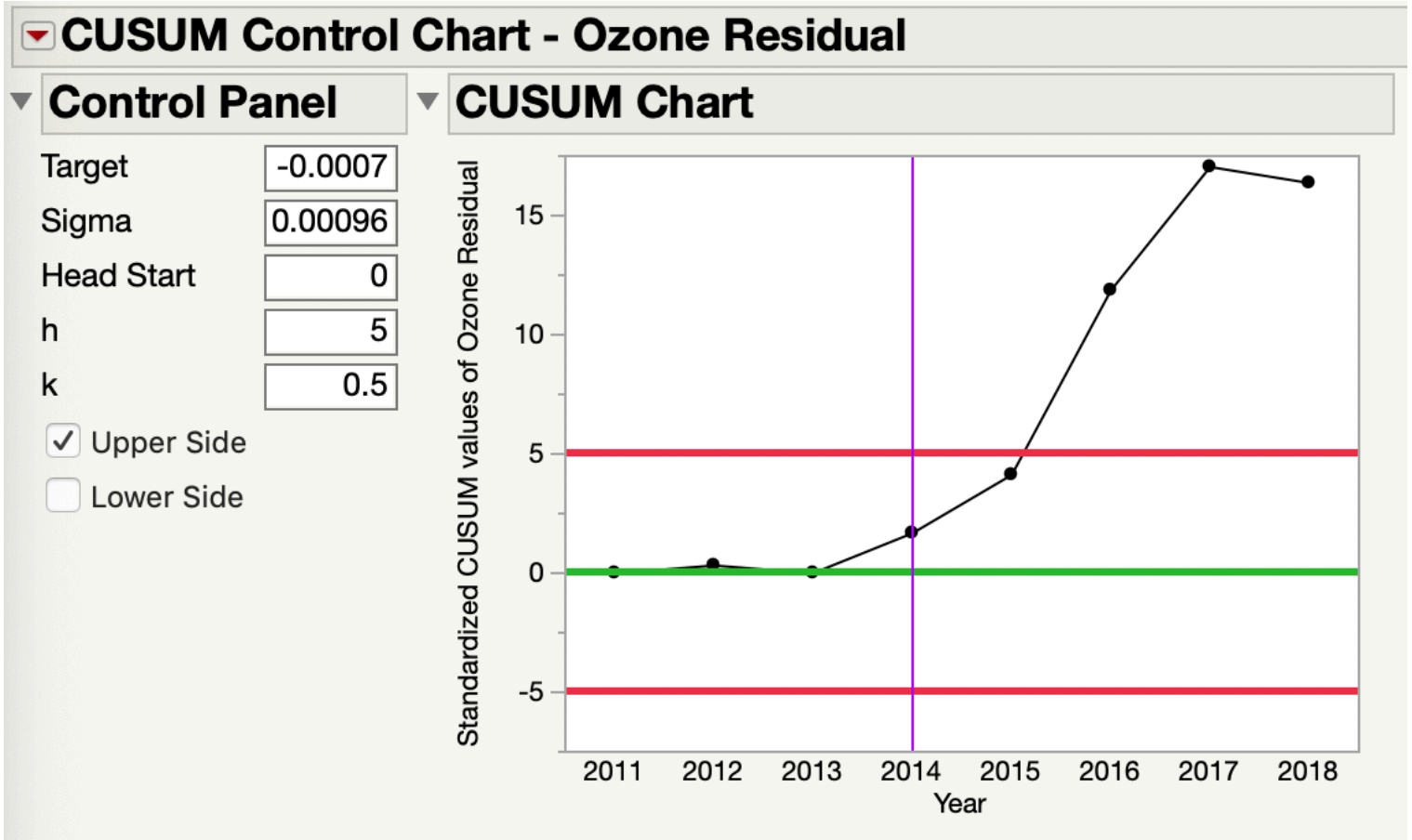
MR Chart



CUSUM Chart - Individual Observation (2011 – 2018)



CUSUM Chart – Averaged over Year (2011 – 2018)



Effect of Ozone

- Ozone is directly related to the Asthmatic conditions.
- An increase in the content of Ozone in the atmosphere, will cause changes in the functions of Lungs and will cause irritation, directly increasing the severity of Asthma.
- Various control exposure studies and few epidemiologic studies also prove this that, an increase in Ozone will have a direct increasing effect on People with Asthma.
- So, Ozone could have been a factor in the significant rise in Average number of patients after 2013, because, we have also established that the Ozone content in Atmosphere has increased around the same time.

Analyzing other factors

01

Similarly, Every other parameters SO2, CO are violating the Independence assumption. So, we use the very same approach.

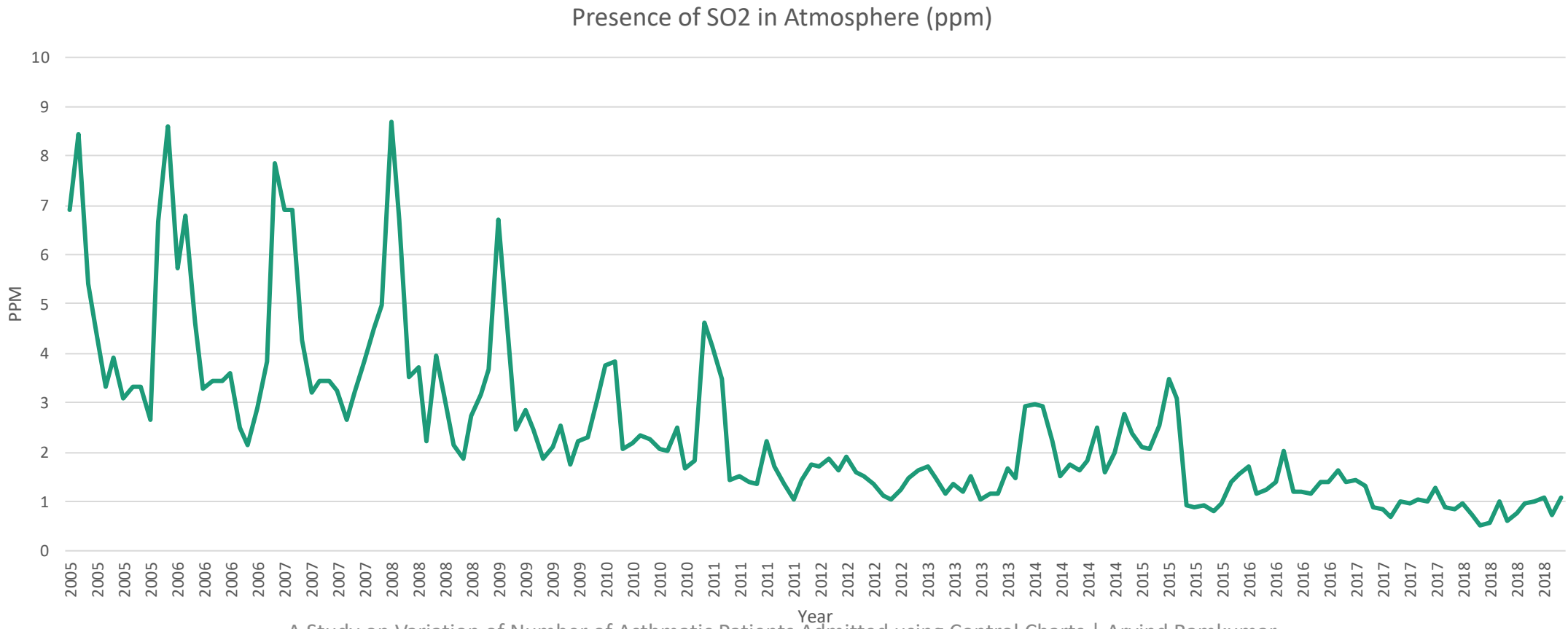
02

We develop a time series model, and we develop the control chart for the residuals.

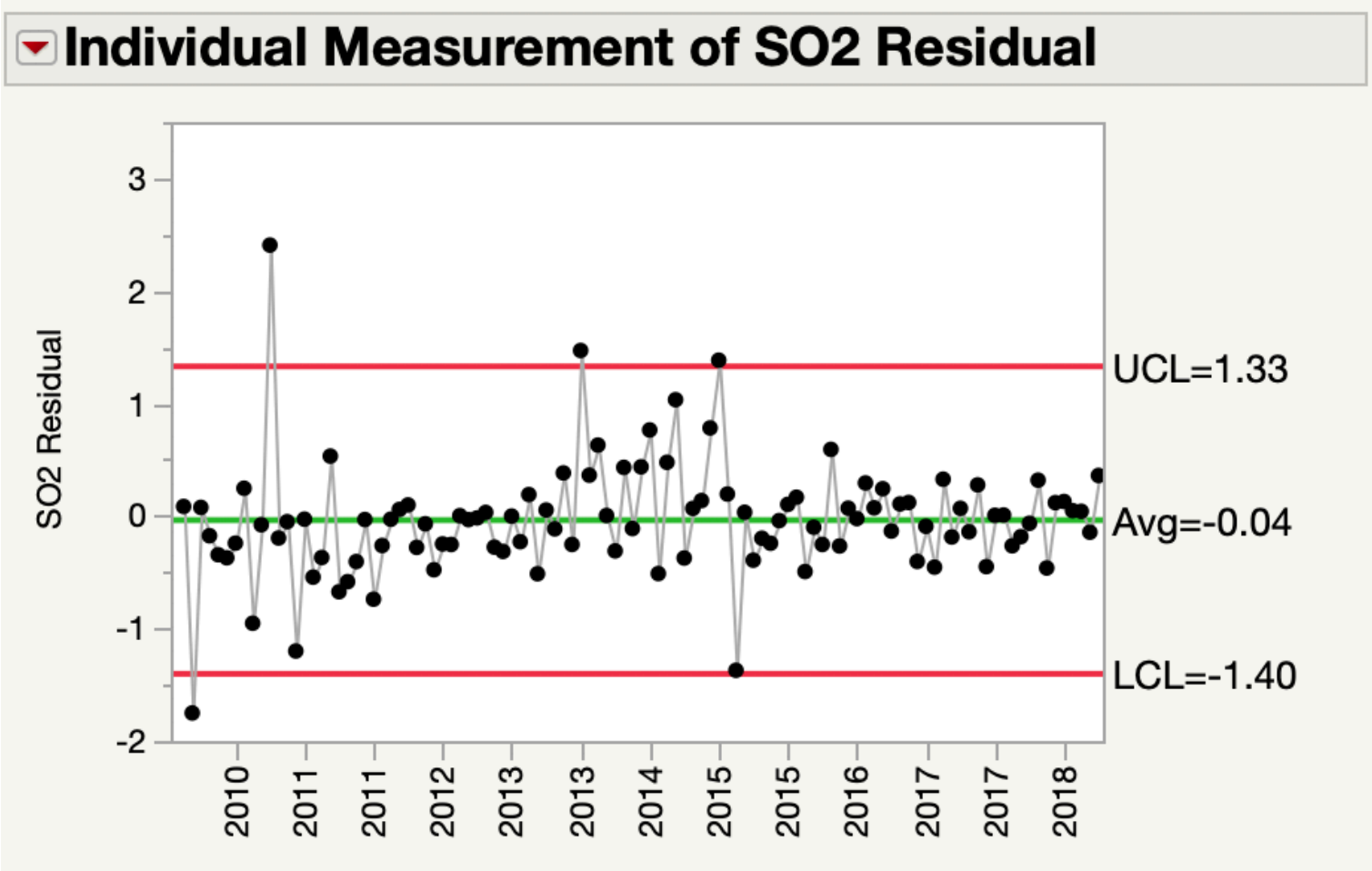
03

We use X/MR Chart initially and because of the very slight variation, we use the CUSUM Chart to understand how it varies.

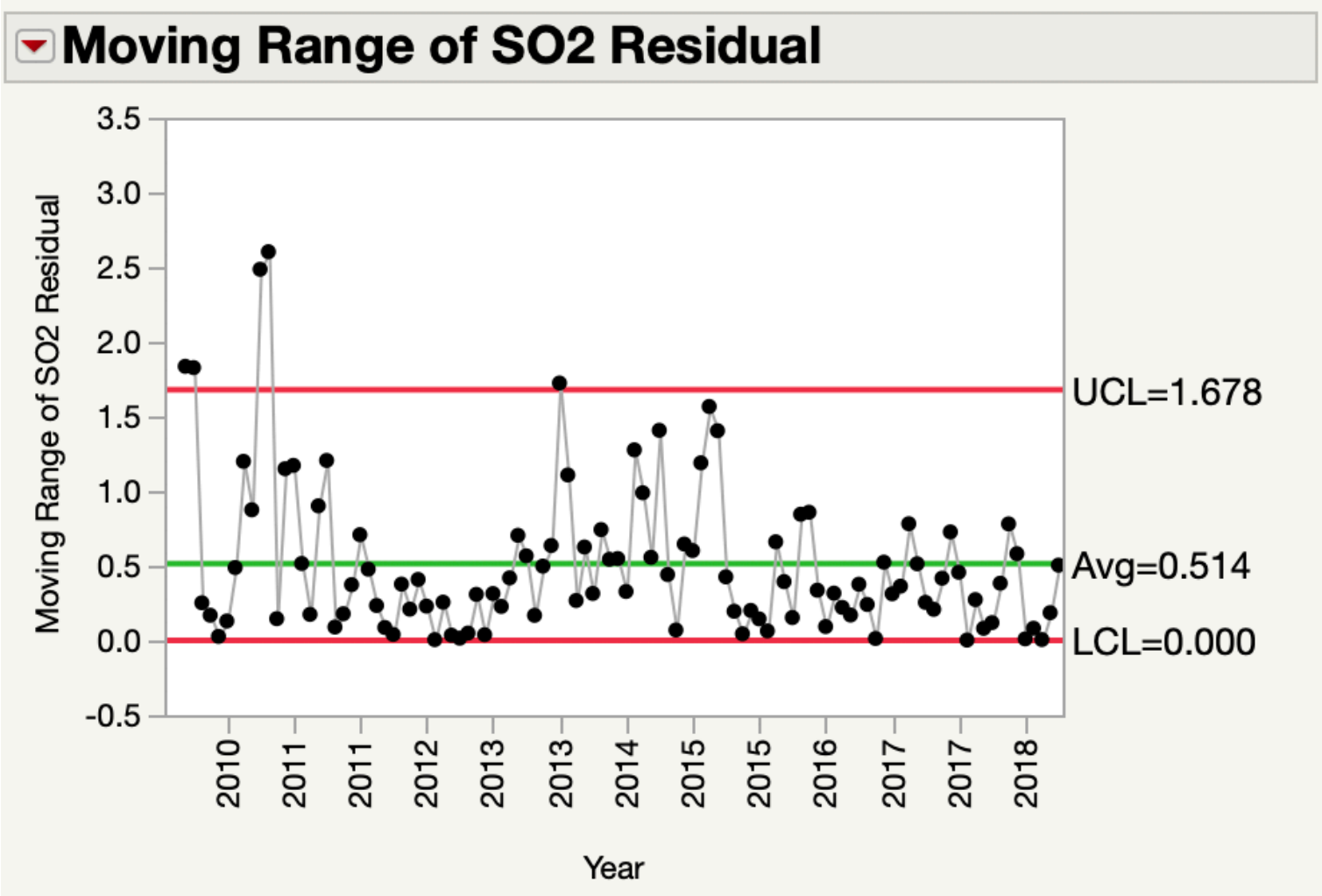
Amount of SO2 in Atmosphere



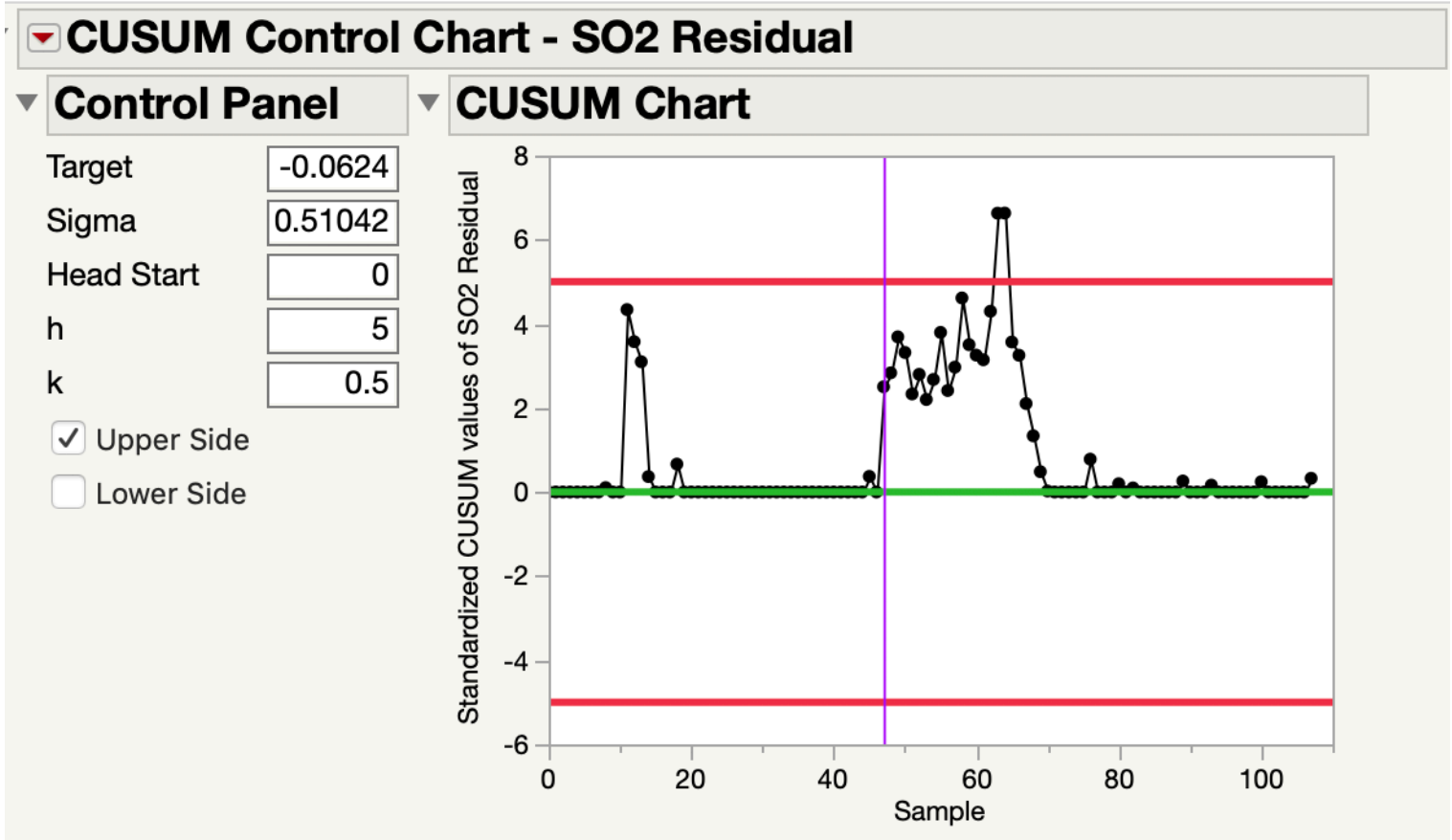
X/I Chart


Individual Measurement of SO2 Residual
SO2 Residual
3
2
1
0
-1
-2
UCL=1.33
Avg=-0.04
LCL=-1.40
2010
2011
2011
2012
2013
2013
2014
2015
2015
2016
2017
2017
2018

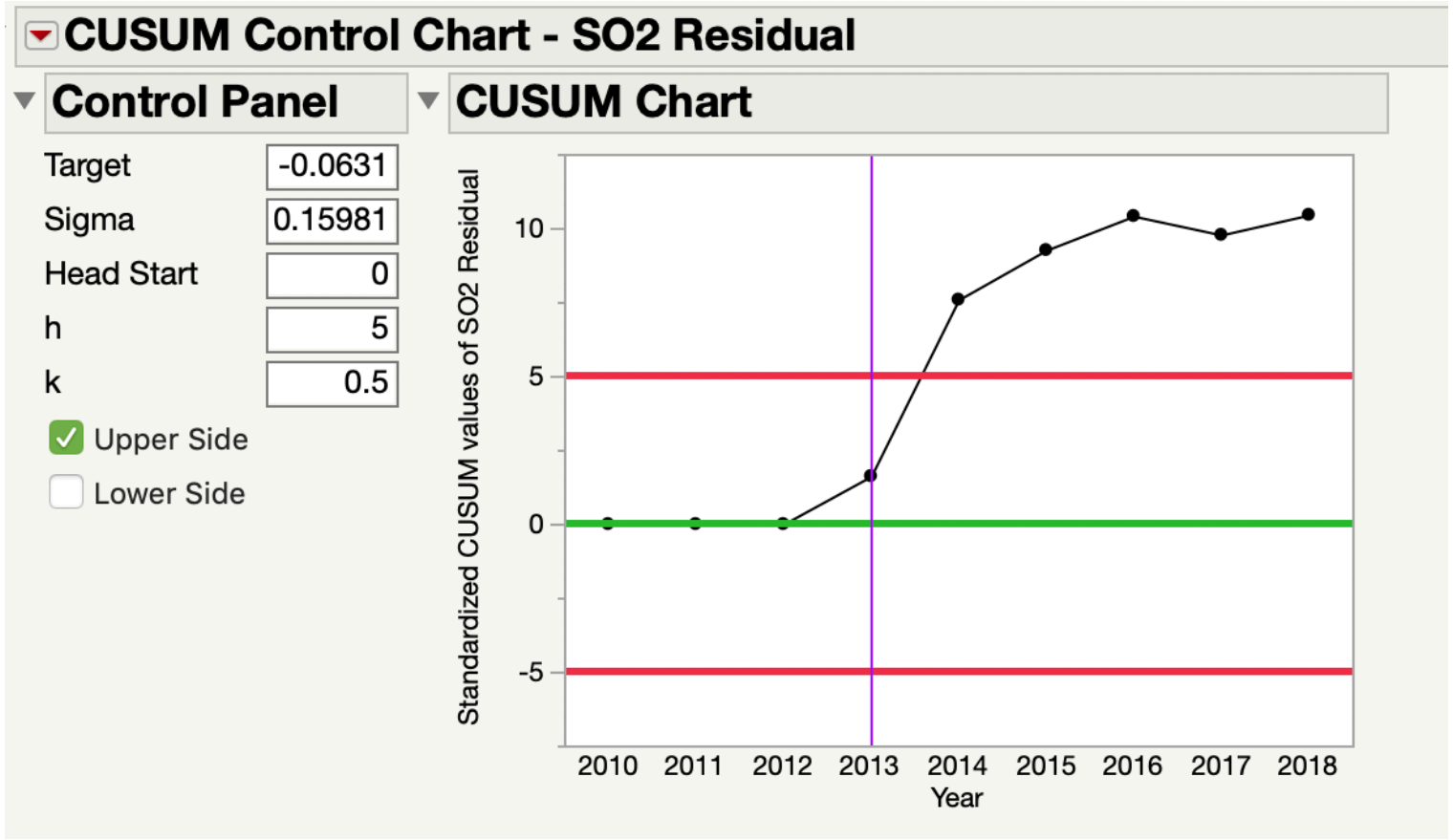
MR Chart


Moving Range of SO2 Residual
Moving Range of SO2 Residual
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
UCL=1.678
Avg=0.514
LCL=0.000
2010
2011
2011
2012
2013
2013
2014
2015
2015
2016
2017
2017
2018
Year

CUSUM Chart – Individual Observation (2010 - 2018



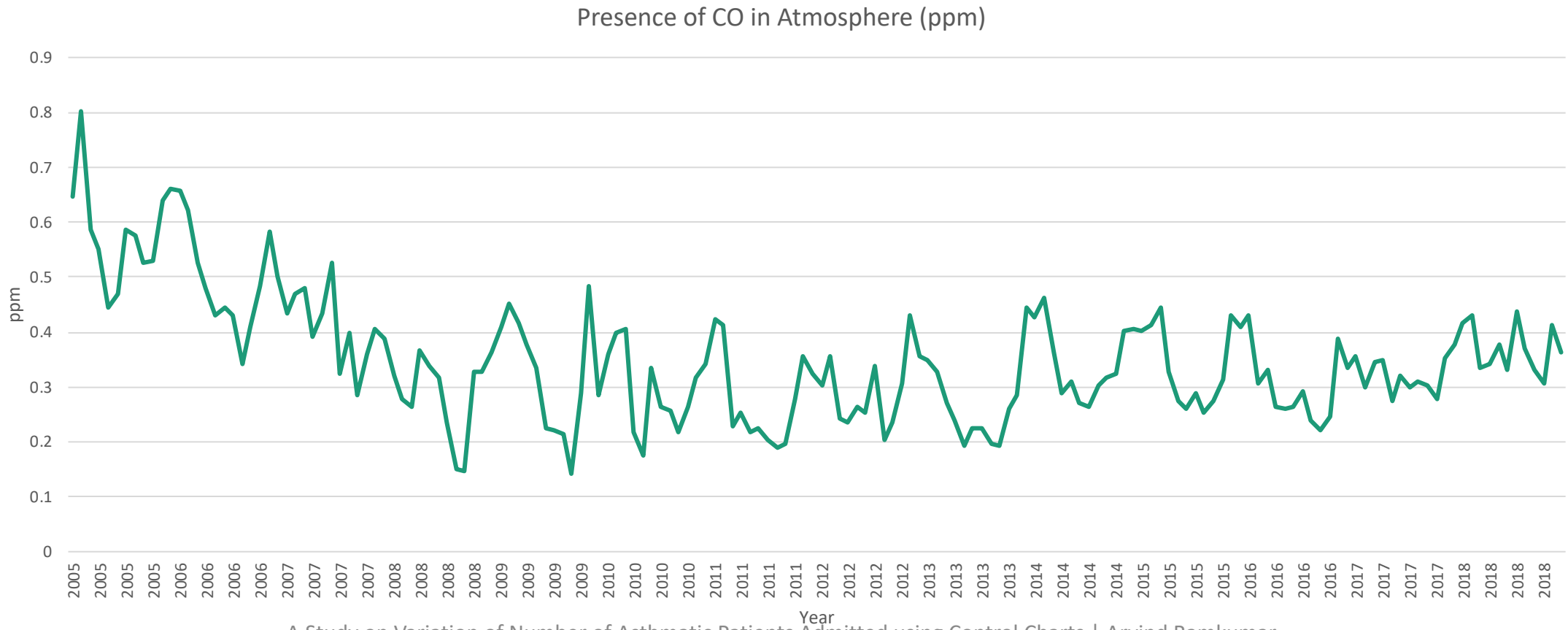
CUSUM Chart – Averaged Over Year (2010 – 2018)



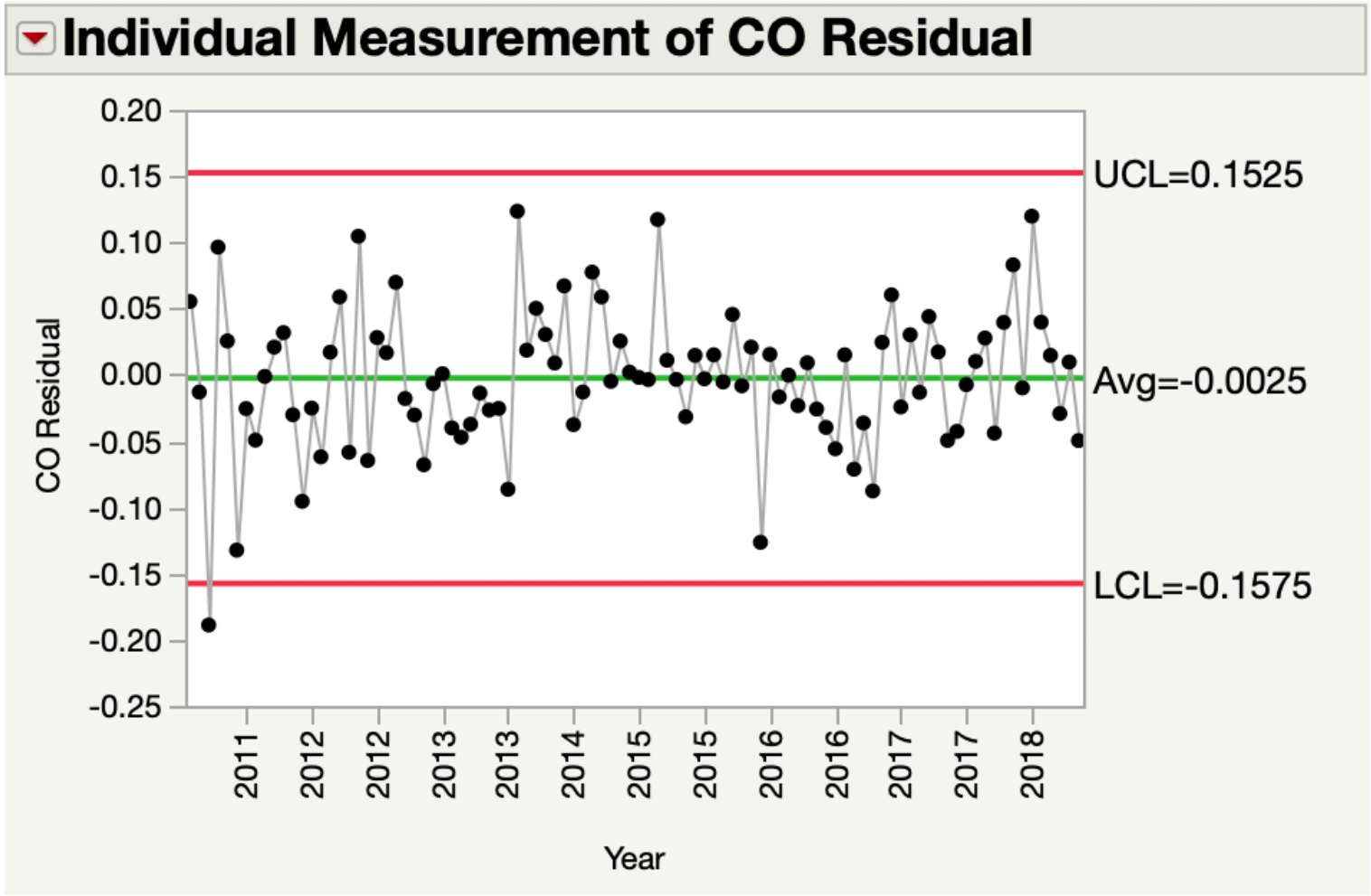
Effect of SO2

- Sulphur dioxide is toxic and is a pollutant. A patient with Asthma are more sensitive to this gas and have a greater inflammatory response.
- Studies prove that even a small trace of SO2 is seen around Asthmatic patients, will have an adverse effect.
- From our study, we realize that there was a sharp rise in SO2 content during the year 2014 and 2015.
- So, there is a very high possibility that it could have had a major impact in increasing the number of patients.

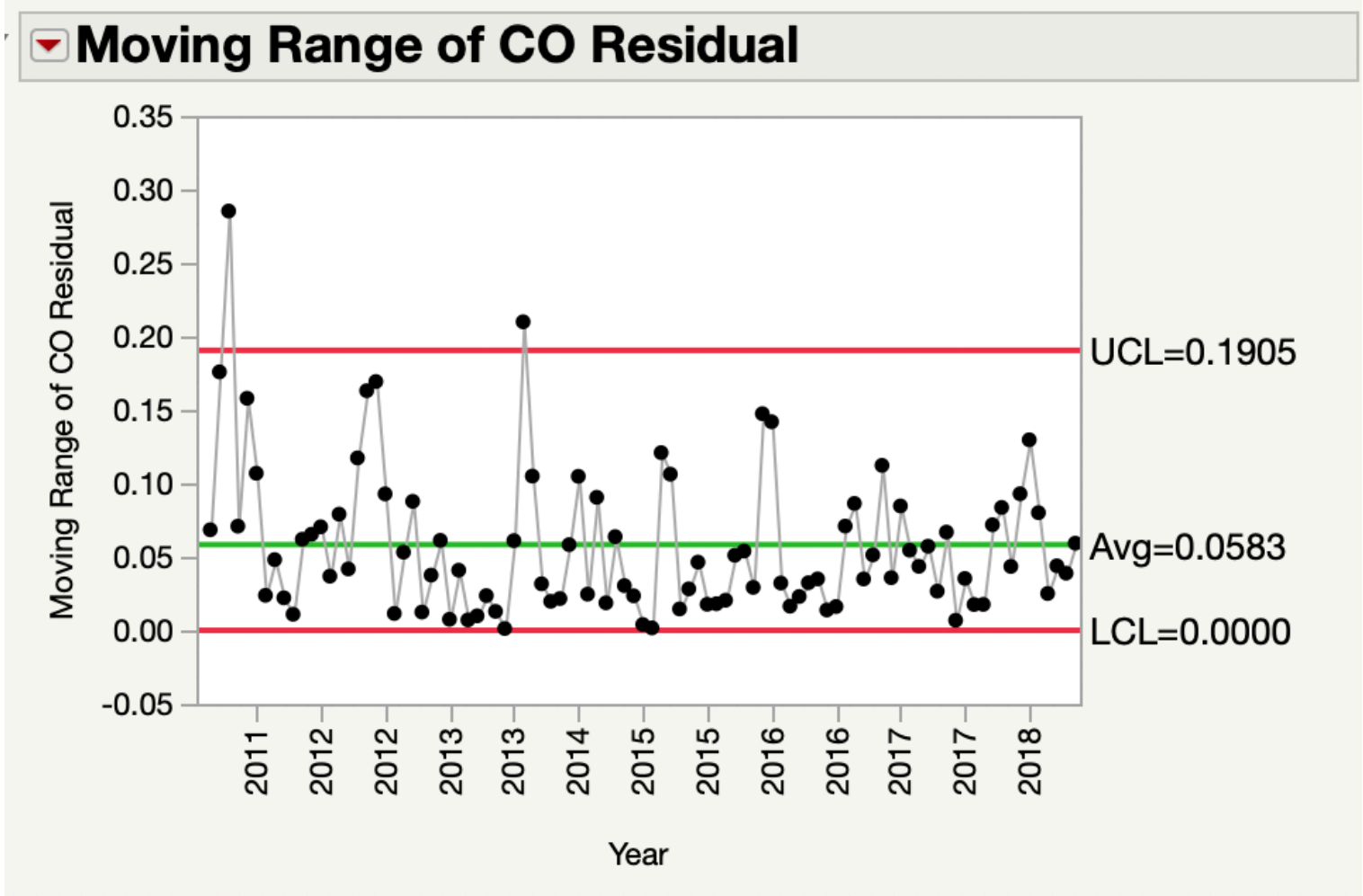
Amount of CO in Atmosphere



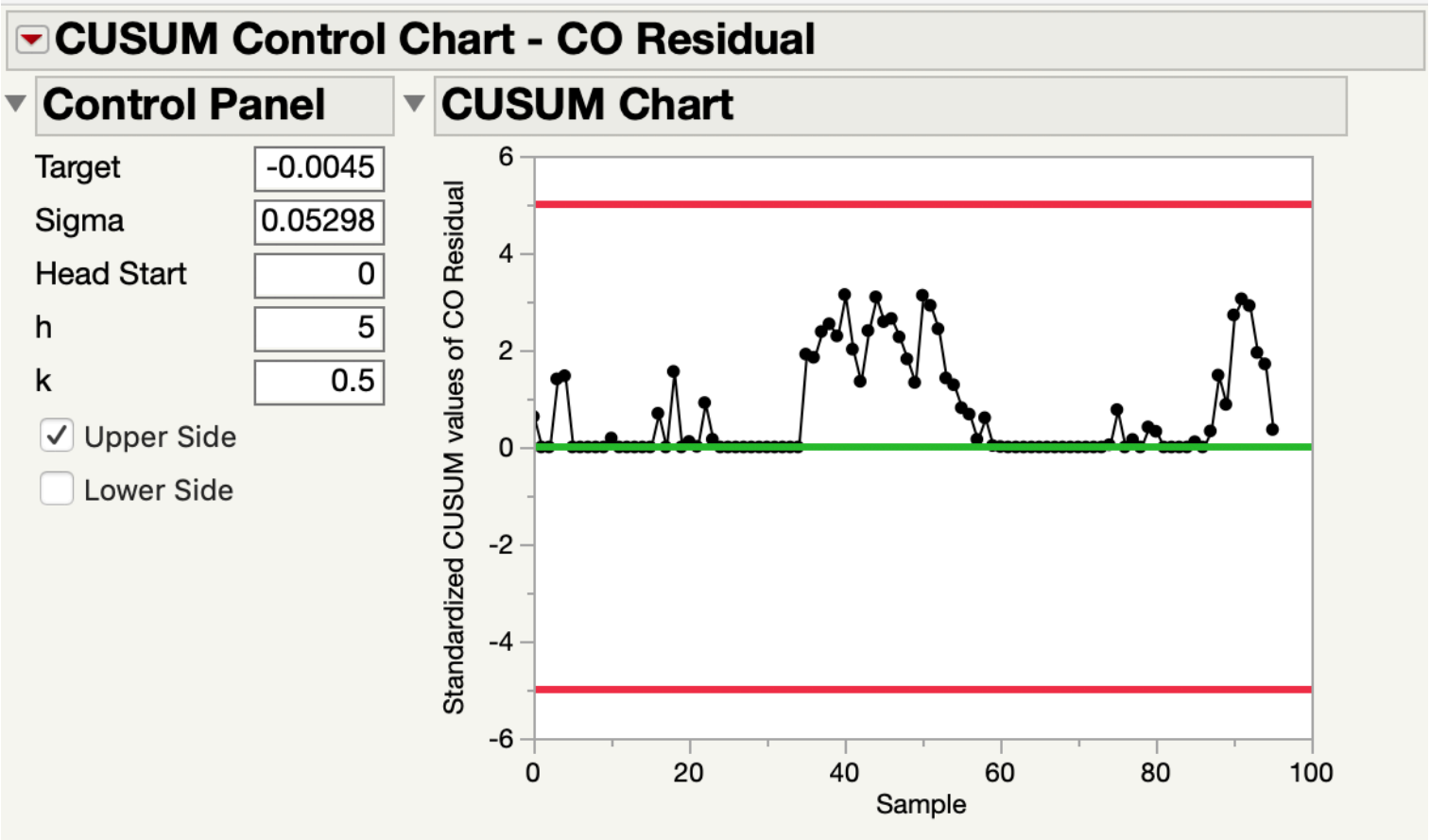
X/I Chart



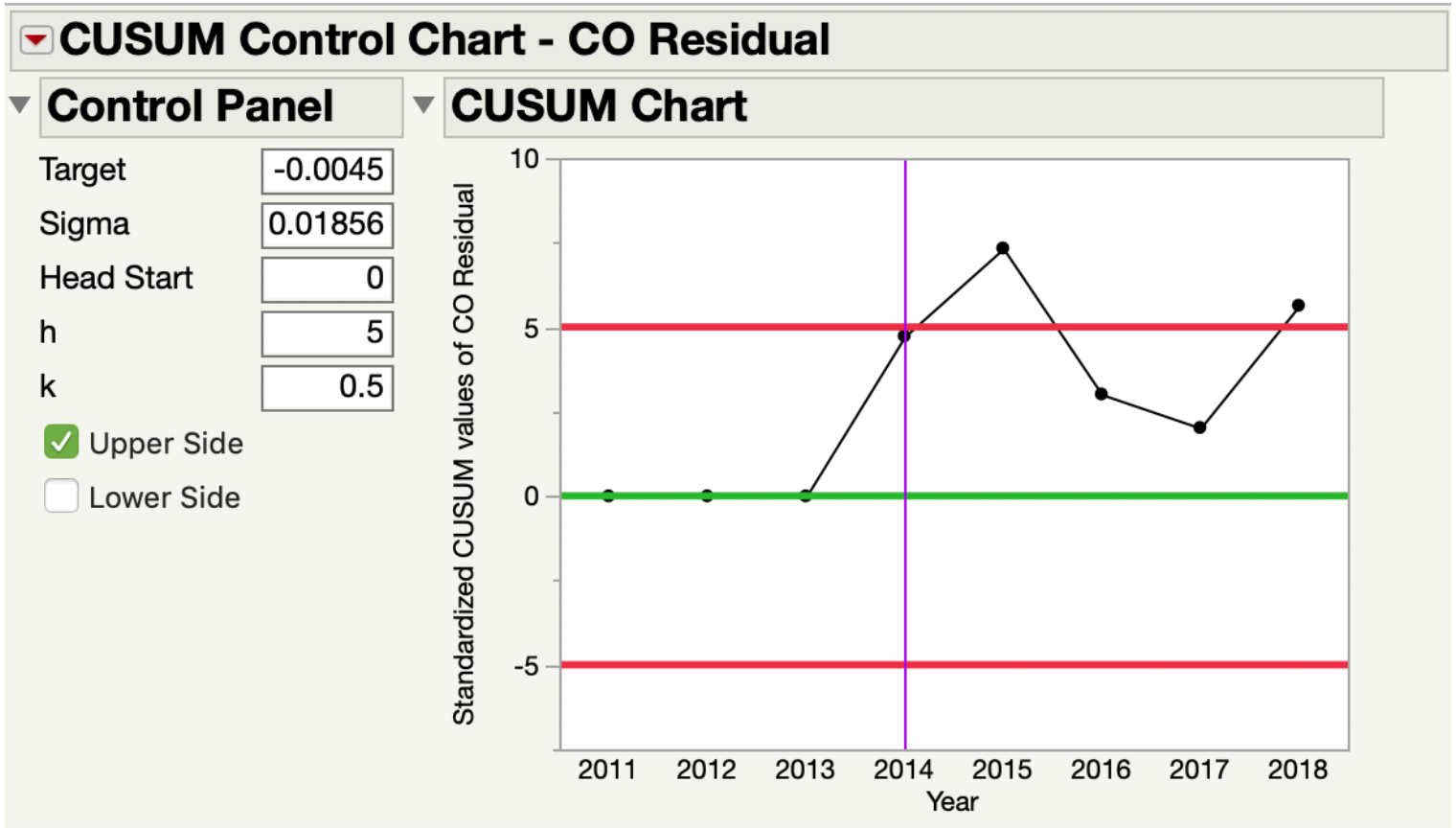
MR Chart


Moving Range of CO Residual
Moving Range of CO Residual
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
UCL=0.1905
Avg=0.0583
LCL=0.0000
2011
2012
2012
2013
2013
2014
2015
2015
2016
2016
2017
2017
2018
Year

CUSUM Chart – Individual Observation (2011 – 2018)



CUSUM Chart – Averaged over Year (2011 – 2018)



Effects of CO

- Although studies are telling us that the effect of CO is much minimal than any other Atmospheric pollutant on Asthmatic patients, these effects can not be ignored.
- From the control charts, the CO content atmoshpere clearly varies at a very high rate.
- Given the minimal effect and the variation, we might want to consider having direct discussion with Doctors and scientists, to make inference based on our study and to understand the effect that it might have on the Increase in the number of Patients.

Future Ideas

A testimony of a Doctor from CHOP (2006), reveals that the other factors including Mold, dust mites, carpeting, bedding etc., can cause asthma effects. So, apart from atmospheric conditions that we have studied, we might get another insight on why there was an increase in patients if we study these factors as well.

Also during a Roundtable conference (2015) about the Climatic conditions with the President, it was quoted that Asthma Rates have been increasing and especially in the lower income community. So, we might want to study the Rise in number of patients based on the income level as well.

It is highly necessary to understand how the number of doctors working in CHOP has changed over the years. This might also be a possible reason of why there was a rise in number of patients.

References

- https://www.haponline.org/Portals/0/docs/Reports-FactSheets/Facts-About-Pennsylvanias-Hospitals-and-Health-SystemsMarch2016.pdf
- https://www.chop.edu/centers-programs/asthma-program/about
- https://www.research.chop.edu/about/research-timeline
- https://www.aafa.org/air-pollution-smog-asthma/