Abstract

Variable Process Control (VPC)

In business, there are many processes or transactions that must be monitored on a periodic basis. Typically these processes are steady-state processes with random daily fluctuation, and can be measured and managed using a Statistical Process Control (SPC) chart.

However, not all processes are steady-state. Some processes exhibit seasonal swings. Other processes even vary by day of week due to shift or delivery schedules. Traditional SPC breaks down under these conditions. This paper addresses a methodology to account for these external forces on the process. I call this methodology Variable Process Control (VPC).

This paper approaches the problem from the standpoint of signal theory. In a measured stream of data, there are two components: signal and noise. Normally, it is the signal that is of interest. However in SPC, it is the noise that we wish to measure and apply the signal after the fact.

In traditional SPC, the signal is the flat-steady state value called the center line. In VPC, the signal varies with respect to time, and must be removed from the measured data stream to leave “flat noise” behind. The flat noise can then be analyzed using standard SPC processes.

Typically, signals have annual or monthly cycles. To handle higher frequency components such as daily peaks and lulls caused by scheduling, the measured data can be considered as a series of distinct, overlapping data series. In other words, a data series for each day of the week.

This paper discusses the theory behind VPC, numerical methods for determining signals and an MS-Excel implementation based on real-world data from processes measured at my current employment.

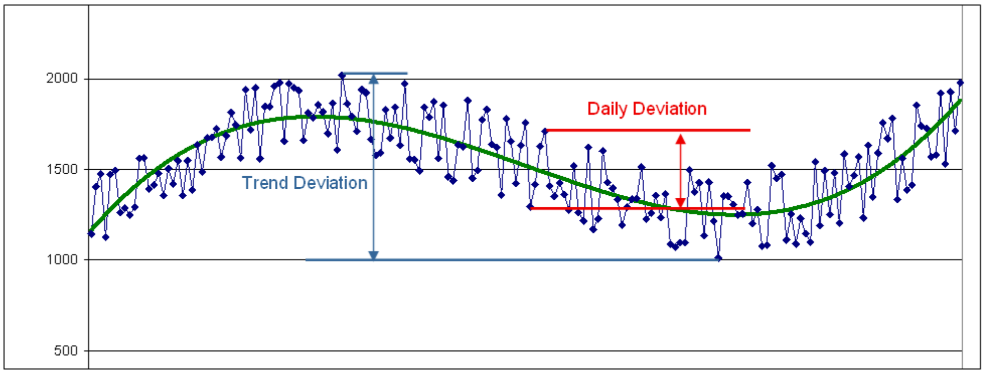
Contents

1 Theory

Measured data are a combination of two forces:

- A long-term growth and seasonal trend (Signal)
- Daily deviations (Noise)

Figure 1- Signal Plus Noise



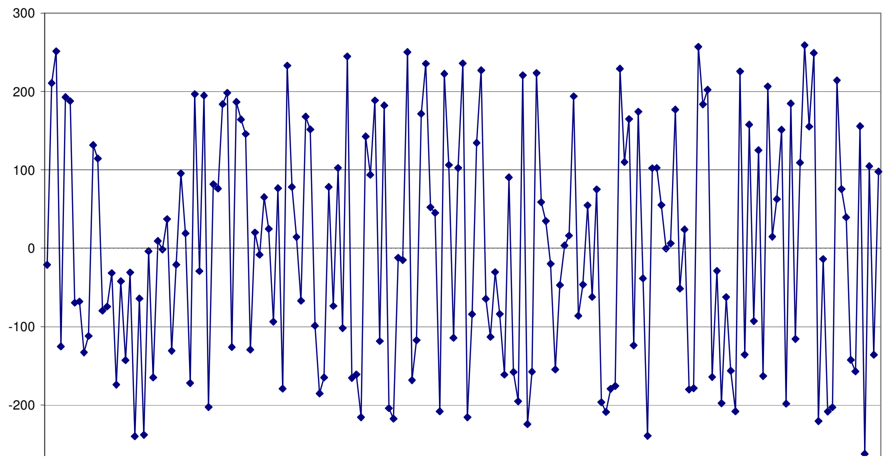
Attempting to apply traditional statistical process control to such data fails.

Applying focus to short-term deviations works only when measuring data along a relatively flat part of the signal. As the signal changes absolute measured values will drift. Applying traditional statistical process control (SPC) to the daily deviations result in false out-of-limit conditions as data rides the curve. For example, a lull relative to its surrounding points in the measured data during peak season may actually be higher than a peak during the lull season.

On the other hand, focusing on long-term deviations ruins the sensitivity of the measurement. Applying statistical process control to the larger trend leads to missed out-of-limits conditions due to the larger seasonal swing. In the data shown above, the daily deviations are typically less than ±300 units from the signal. However, over the course of the entire cycle, the overall deviation is ±1,000 units. Applying the larger deviation for the long-term fluctuations means that while a value is within limits for the entire cycle, it is well out of range for its expected daily value.

The solution comes from "flattening" the deviations. This is accomplished by computing the value of the signal and mathematically subtracting it from the measured data.

Figure 2 - Flattened Deviations



The chart shown in Figure 2 is the same data as shown in Figure 1 with the signal removed. This data has the seasonality removed, and can be subjected to traditional SPC.

1.1 How Much Data is Enough?

The amount of data collected depends on what you want to do with it. The main objective of this development was to come up with a process to evaluate recent transactions to see if the number of transactions is consistent with expectations. A different mindset may apply if evaluation is done for the purposes of forecasting.

The advantage of tracking a large amount of historical data is that it produces a very steady trend line. While this is useful for forecasting, it may not be suitable for tracking processes where the process itself is subject to change. Having the trends rooted in a large data set means that they are almost fixed and sluggish to respond to more recent changes.

On the other hand, considering too little amount of data leads to an unstable trend line that varies too much and may not be indicative of the process.

Most of the data analyzed in this study had a period of a year, and by a process of repeated evaluation, a half cycle (180 days) worth of data was deemed the best compromise.

2 Computing the Signal

Getting a good curve fit to the data is essential to the process. A process of least-squares fit was used to best fit an equation to the measured data.

There are two methods for computing the signal:

- General – The general model works with almost any kind of a curve.
- Special – The special model works with data that is a half cycle in span or less.

Before engaging either method, remove any obvious out-of-limit data from the sample. These may show up as single spikes or they could be shown as a sudden shift in volume that levels off to define a new trend.

2.1 General Method

The General Method has the advantage in that it can handle any time range, is truly cyclic and could be used to forecast. It has the disadvantage of being cumbersome to set up, and requires human intervention to compute.

Most processes consist of a steady state component, a growth component, and a seasonal component.

- The steady state component is the level at which the measured data tends to settle. In the traditional SPC model this is the center line.
- The growth component is the rate of expansion or decline typically calculated on an annual basis. Usually, this is an exponential expression, however, for small growth rates, and relatively short periods (several years vs. decades) the expression can be accurately depicted by linear growth.
- The seasonal component is periodic. This can be represented by a sine equation.

The general equation then becomes:

A*t + B + C * Sin (t / D + E)

Where:

- A is the linear growth component
- B is the steady-state component
- C is the amplitude of the seasonal component
- D is the period of the seasonal component
- E is the phase shift of the seasonal component (when the cycle begins)

There is no easy algebraic solution to this equation. The method this paper discusses uses MS-Excel's solver feature to compute the least squares fit of the parameters. Since Excel assumes radians as the unit for trigonometric functions a 2 * PI() should be added to the equation making it:

A*t + B + C * Sin (2 * PI() * t / D + E)

A side discussion on t is in order. Time, as measured in Excel, is the number of days since January 1, 1900. For example, July 6th 2009 is 40,000 days since the beginning of the epoch. There is a concern that using such a large number in the background calculations will cause overflow and possible rounding errors. So an Elapsed Date (E-Date) is used in the calculations. The E-date merely moves the beginning of the epoch to a more recent baseline date such as January 1, 2009.

The E-date is calculated by subtracting the baseline from the current date:

E-Date = Observed Date – Baseline

The E-Date assures that we are working with numbers that are in the hundreds instead of tens of thousands. When these numbers are raised to higher powers, the order of magnitude makes a difference. It does not matter if the E-Date is a negative number.

2.1.1 Setting up Solver

Setting the Solver Utility is essential to computing the coefficients of the least squares fit signal.

First arrange the data on the spreadsheet as follows:

Figure 3 – General Method Spreadsheet Setup

	A	B	C	D	E	F
1	**Date**	**Value**	**E-Date**	**Signal**	**Delta^2**	**Adjusted**
2	05/03/09	23960	-92	22960.08	999833.3	999.9167
3	05/04/09	12623	-91	11623.01	999981.5	999.9907
4	05/05/09	12485	-90	11485.23	999537.1	999.7685
5	05/06/09	8476	-89	7476.75	998500.6	999.25
6	05/07/09	11052	-88	10053.56	996873.3	998.4354

	I	J	K	L	M
1	**A**	**B**	**C**	**D**	**E**
2	1	0	1000	365	0
3					
4	**Minimize**	85887118			

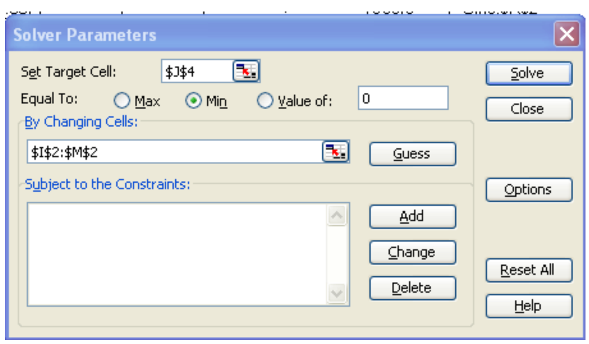
Named ranges are used to make the formulas easier to read and to make it easier to visualize what is going on.

Table 1 - General Method Spreadsheet Setup

Column / Cell	Named Range	Value / Formula	Comments
A	N/A	The measured date	
B	N/A	The measured value	
C	EDate	=A2-Baseline	Baseline is a named range defined as the value 1/1/2009
D	Signal	=CoefA*C2+CoefB+CoefC*SIN(2*PI()*C2/CoefD + CoefE)	Signal Equation
E	Delta2	=(B2-D2)^2	The difference between the observed value and the calculated value squared..
F	Noise	=B2-D2	The measured value with the signal taken out of it.
I2	CoefA		The growth component coefficient
J2	CoefB		The steady-state component coefficient
K2	CoefC		The amplitude of the seasonality
L2	CoefD		The period of the seasonality
M2	CoefE		The phase shift of the seasonality
J4	N/A	=SUM(Delta2)	The sum of the squares of the deviations. It is the sum of the squares that is to be minimized

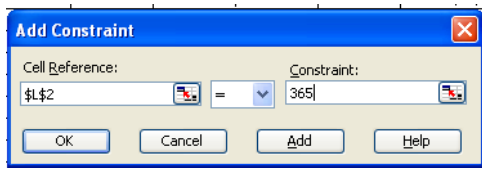
The initial coefficients are set up by educated guesswork. A visual inspection of the plotted data should be sufficient. The only coefficient that has to be set carefully is the period. Normally, this is set to a year (365 days). However, other minor cycles such as monthly or even weekly cycles may exist. If set far enough away from the observed cycles, solver may latch onto one of these minor cycles.

In Excel 2003, select Tools → Solver, and set up the Solver solution as shown below:



Click on Solve. The process will run and you will get a first approximation of the coefficients. Keep re-running solver until the value in Cell J4 is constant.

As an optional step, if it is known that data actually conforms to a known period, this value can be locked in the solver calculations. Click on the Add button next to the “Subject to Constraints” box and add in the constraint.



In this case, the period is forced to be one year. The rest of the coefficients will be recalculated to best fit a one year period.

For the 80+ data sets under study, those that had periods had a period that computed to with 10 days of 365. I elected to lock in the period at 365 days. Data sets with periods well away from 365 days are handled by a different methodology in the analysis.

2.2 Specific Method

The specific method has the advantage of being easy to implement, is totally algebraic and requires no human intervention. It works well for time ranges of a half a period; for example last 180 days worth of data of an annual cycle. It is not guaranteed to work well with a time range greater than a half cycle, nor is it recommended for forecasting. However, if tracking recent rolling history, such as the last 180 days of performance, it is an ideal method.

The specific method is based on approximating the general equation with a third degree polynomial. In scores of data sets, the polynomial overlaid the general equation perfectly.

Setting up the specific method requires a slightly different spreadsheet layout

Figure 4 - Specific Method Spreadsheet Setup

	A	B	C	D	E	F	G	H	I
1	**Date**	**Value**	**E-Date**	**Signal**	**Noise**	**DOW**			
2	05/03/09	23,960	122	11,754	12,206	Sun		**A**	0.000658
3	05/04/09	12,623	123	11,786	837	Mon		**B**	-0.495618
4	05/05/09	12,485	124	11,818	667	Tue		**C**	124.223
5	05/06/09	8,476	125	11,850	-3,374	Wed		**D**	2780.044

Table 2 - Specific Method Spreadsheet Setup

Column / Cell	**Named Range**	**Value / Formula**	**Comments**
A	Measured Date	N/A	
B	Measured Value	N/A	
C	EDate	= A2-Baseline	Baseline is a named range defined as the value 1/1/2009
D	Signal	= CoefA*C2^3 + CoefB*C2^2 + CoefC*C2 + CoefD	Third degree polynomial.
E	Noise	= B2-D2	The measured value with the signal taken out of it.
F	DOW	= TEXT(A2,"ddd")	The day of the week of the measured data.
I2	CoefA	See below	
I3	CoefB	See below	
I4	CoefC	See below	
I5	CoefD	See below	
I2:I5	N/A	=cubest(EDate,Measured)	This formula is entered as an array formula. The cubest user-defined function is defined in the appendix

3 Using the Computations.

A plot of the measured data and trend line reveals that there are three types of control limits:

- General
- By Day of Week
- Scattered

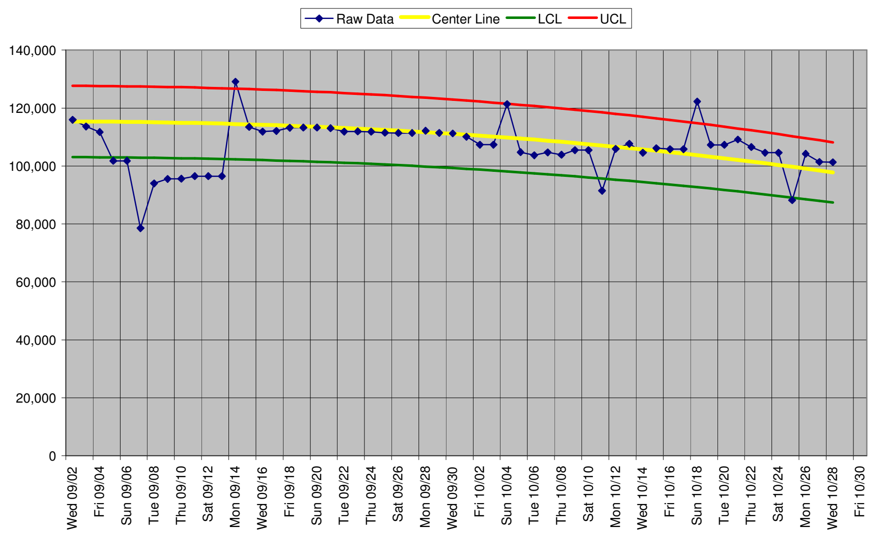
3.1 Types of Control Limit Charts

Looking at the graphs reveals its type, but assigning the control limit does not have to be a matter of subjectivity. Determining the type of control limit to use can be calculated automatically by user-defined parameters.

In the following graphs, four pieces of data are presented:

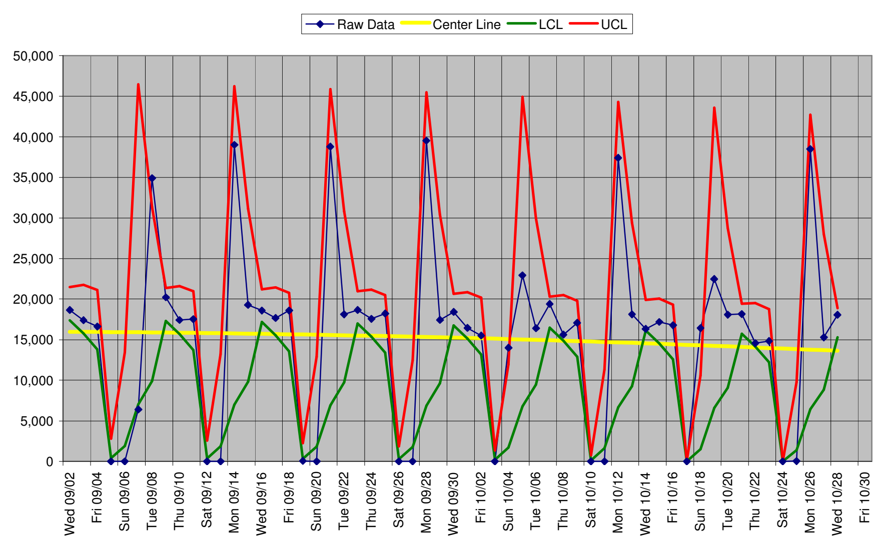
- The measured value (blue line with diamonds)
- The center line or signal (solid yellow line)
- The lower control limit (solid green line)
- The upper control limit (solid red line)

Figure 5 - General Type Control Limit Chart



General charts wrap themselves around the trend line rather closely. Notice how the signal and control limits follow the data.

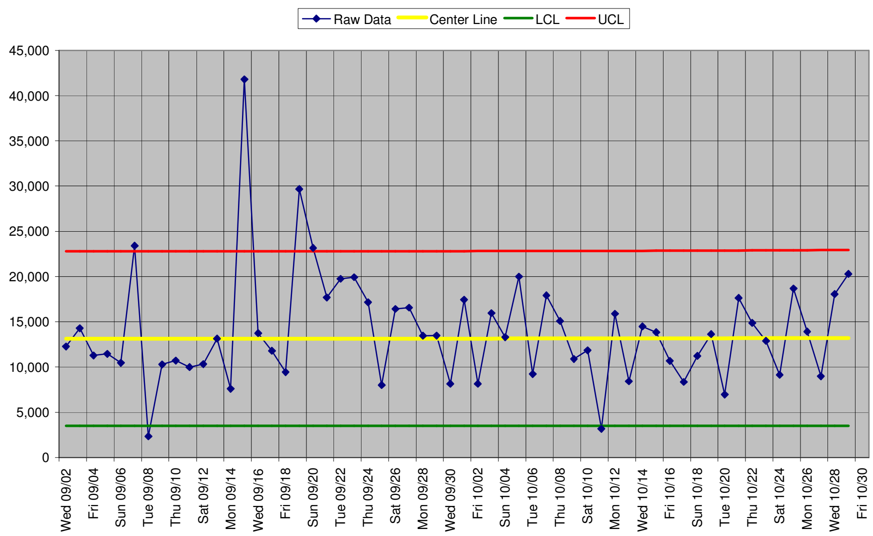
Figure 6 - By Day of the Week Control Limit Chart



By Day of the Week charts have a very distinct pattern that varies by day of the week. In this particular instance, the process is influenced by the following factors: A steady stream of deliveries are made on all days of the week. However, no processing is done on Saturday or Sunday. Monday is an extra heavy day of production to deplete the weekend buildup. The rest of the week's processing is dedicated to removing residual backlog and current daily shipments.

This type of process is essentially 7 separate processes and will be treated as such.

Figure 7 - Scattered Control Limit Chart



Scattered Control Limit charts are similar to general charts except that the data are more widely scattered from the centerline. In this case, the center line (signal) is ignored and hard value control limits are used. These charts typically represent two types of processes:

- Erratic: the daily data fluctuations are large.
- Sporadic: days or weeks may go by between transactions (daily transaction volume = 0) with brief periods of high volume activity.

3.2 Decision Making Parameters

This part of the paper describes the process known as the Control Limit Calculator.

For the purposes of the evaluation upon which this paper is based, it was deemed that transactions that exceed 2-sigma deviations should be flagged. Furthermore, rather than expressing the deviation as fixed values the deviations are evaluated as percentages of the signal or varying centerline value. Therefore, the control limits will "follow" the center line as it varies.

This brings into play the rest of the spreadsheet.

Figure 8 - Control Limit Calculations

	H	I	J	K	L	M	N	O	P
7		Overall	Sun	Mon	Tue	Wed	Thu	Fri	Sat
8	Average	12,903	15,181	12,237	12,748	13,636	12,404	10,348	13,636
9	Stdev	4,737	6,241	4,477	4,153	3,966	2,565	4,902	5,008
10	Percent	73%	82%	73%	65%	58%	41%	95%	73%
11	Avg from TL		2300.926	-657.5862	-160.3011	730.5998	-497.9815	-2561.383	718.0532
12									
13	5th %tile	6451		6451	22108				
14	95 %tile	22108							
15									
16	Use Metric	1		Override					

All three of the control limit types are evaluated here. The general principle is that for each series of data (overall data if a general type, or 7 distinct series if by day of the week) the percent from the trend line is given as:

2 * Sigma / Average (Measured Data).

These are the percentages that appear in row 10 of

Figure 8.

For any set of data, all three types of control limits are calculated, and the best set is picked. In addition, some adjustments are made for known anomalies in the process. These anomalies will be pointed out as the formulas are discussed:

3.2.1 Computing General Control Limits.

The computation of the general control limit is simple and straightforward. The relevant cells are:

Cell	Formula
I8	=AVERAGE(Measured)
I9	=STDEV(Noise)
I10	=IF(I8>0,2*I9/I8,0)

3.2.2 Computing by Day of the Week Control Limits

The computation of by Day of the Week Control Limits is similar to general control limits except it uses array formulas instead of regular formulas. This is where the DOW column (Column F) comes into play. The computation for each of the days of the week are similar, so only Sunday's computation is shown in detail.

Cell	Formula	Comment
K8	=AVERAGE(IF(DOW=J7,Measured,""))	If the day of the week matches the day of interest, then average the measured values. Ignore measured values for other days of the week.
K9	=STDEV(IF(DOW=J7,Noise,""))	If the day of the week matches the day of interest, then take the standard deviation of the measured values. Ignore measured values for other days of the week.
K10	=IF(J8>0, IF(SUMIF(DOW,J7,Measured)/SUM(Measured)<MaxVol, MIN(2*J9/J8,MaxPCT), 2*J9/J8), MaxPCT)	Some heuristics are applied here. See the explanation at the bottom of the table.
K11	=AVERAGE(IF(DOW=J7,Noise,""))	This formula measures the average deviation for the day of the week form a general trend line. This value is used to plot and compute the control limits.

The standard process is that shipments occur on all days of the week, but stack up on weekends and the surplus is processed on Monday. There are exceptions to the process. Occasionally week end processing occurs. This leads to low-volume, high-deviation metrics.

To account for the occasions where this might occur, an addition check is performed: Compute the deviation for the day of the week. If the total volume for the day of the week is less than 10% (MaxVol) of the weekly volume, then assign either 75% (MaxPct) or the computed volume, whichever is less.

The figures 10% and 75% were reached by looking at the number of times when off-day processing occurs, and the tolerance for deviation. Other values could have been assigned based on what the history indicates. These values work well for the 80+ processes evaluated in this study.

This is one way of addressing these deviations. It will allow for some processing on these days without flagging them. The other option is to not consider processing on these days at all (set the 2-Sigma deviation to 0% regardless of the history).

3.2.3 Computing Scattered Control Limits

There are instances where neither a general control limit calculation nor by day of the week calculation suffices. In this instance the control limits are set by percentile. The lower control limit is the 5^{th} percentile and the upper control limit is the 95^{th} percentile. These percentiles are modified under the following conditions:

- If 10% or more of the measured values are zero, the lower control limit is zero
- If there is no data for 10% or more of the available dates, the lower control limit is zero.
- If the lower control limit is zero, then the upper control limit is the 95^{th} percentile of all non-zero values.

The percentile method was chosen because in these cases the general model fails. The general model works on a percentage of the signal for a given measured time. Scattered data often have signals with very bizarre characteristics. For example if there is a lot of activity at the beginning of the period, and none at the end of the period, the trend line will plunge until it hits the X-axis. Or, the standard deviation is well in excess of $\pm$100%.

Cell	Formula	Comment
I13	=ROUND(IF(OR(COUNTIF(Measured,0)/COUNT(Measured)>0.1, COUNT(EDate)/(MAX(EDate)-MIN(EDate))<0.1), 0, PERCENTILE(Measured,0.05)), RoundValue)	Round value is zero for volumes and 4 places for when the measured data is percentages instead of whole numbers.
I14	=ROUND(IF(I13=0, PERCENTILE(IF(Measured>0,Measured,""), 0.95), PERCENTILE(Measured,0.95)), RoundValue)	

3.2.4 Determining the Best Fit

Determining the best fit is a matter of comparing the values for each of the methods to one another.

If each of the by Day of the Week percentages is less than the general percent, and each percent is also less than or equal to75%, then the Day of the Week is a better fit for the data than the general percent.

If the first test fails and the general percent is less than or equal to 75%, then the general percent is the best fit.

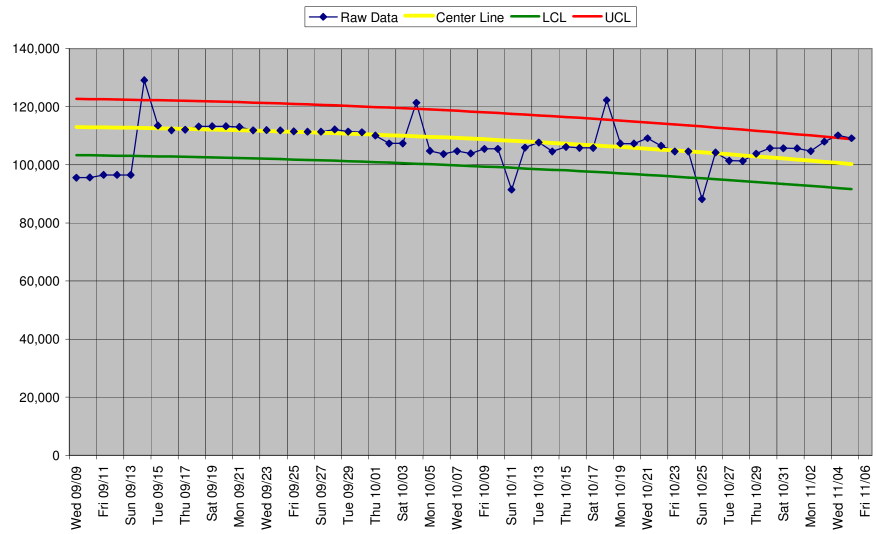
If both of the above tests fail, then the scattered control limits are the best fit.

In practice, these tests work out very well. The type of control limit predicted by the tests do, indeed, match the corresponding chart.

4 Applying the Control Limits

Control limits are measured against the trend line. In normal statistical process control this is a horizontal line representing the average of the measurements. In variable process control, this is the signal.

Figure 9 -General Percent Control Chart



The spreadsheet that applies the control limit needs the parameters computed in the control limit calculator. These include the coefficients of the trend line formula (for this example, the coefficients of the 3rd degree polynomial).

Figure 10 - Coefficients for 3rd Degree Polynomial

	L	M
2	A	-0.004033
3	B	-2.25441
4	C	114.0574
5	D	111950.5

Other parameter include the 2-sigma deviations (Cells O2:U2) for day of the week type controls, and the average deviation from general center line (Cells O3:U3) for day of the week type controls.

Figure 11 - Standard Deviations and Deviation from General Center Line

	O	P	Q	R	S	T	U
2	Sun	Mon	Tue	Wed	Thu	Fri	Sat
3	0.085764	0	0	0	0	0	0
4	0	0	0	0	0	0	0

Cells O2 and P2 serve "triple duty". If the type of control is general percent, then the general percent is contained in cell O2. If the type of control is percentile, then cell O2 is the lower control limit and cell P2 is the upper control limit.

Figure 12 – Sample Computations for General Percent Control

	A	B	C	D	E	F	G	H
1	Date	Value	E-Date	Center Line	DOW	New CL	LCL	UCL
2	10/11/09	91,442	67	108,259	Sun	108,259	98,975	117,544
3	10/12/09	105,929	68	108,014	Mon	108,014	98,750	117,278
4	10/13/09	107,643	69	107,762	Tue	107,762	98,520	117,005
5	10/14/09	104,577	70	107,505	Wed	107,505	98,285	116,725
6	10/15/09	106,139	71	107,241	Thu	107,241	98,043	116,438
7	10/16/09	105,820	72	106,971	Fri	106,971	97,796	116,145
8	10/17/09	105,820	73	106,694	Sat	106,694	97,544	115,845
9	10/18/09	122,241	74	106,411	Sun	106,411	97,285	115,538

Column	Formula	Comment
A	N/A	Measured Date
B	N/A	Measured Value
C	=IF(AND(ISNUMBER(A2),A2>0),A2-Baseline,"")	The elapsed date (E-Date)
D	=MAX(IF(ISNUMBER(B2), CuA*C2^3+CuB*C2^2+CuC*C2+CuD, NA()), 0)	Center line value computed using the coefficients from the control limit calculator (Figure 10 - Coefficients for 3rd Degree Polynomial) The center line is capped at a minimum of zero.
E	=TEXT(A2,"ddd")	Day of the week. This value is used to compute a by day of the week trend line and is otherwise ignored.
F	=D2+HLOOKUP(E2,MetDOW,3,FALSE)	Day of the week trend line. Each day of the week has an average deviation from the general trend line. This formula computes a modified trend line from the general trend line for each day of the week essentially splitting it into 7 separate series. See Figure 11 - Standard Deviations and Deviation from General Center Line.)

Column	Formula	Comment
G	=MAX(IF(ISNUMBER(A2), IF(MetricType=1,D2-D2*HLOOKUP("Sun",MetDOW,2,FALSE), IF(MetricType=2,HLOOKUP("Sun",MetDOW,2,FALSE), IF(MetricType=3,F2-F2*HLOOKUP(E2,MetDOW,2,FALSE),NA()))),NA()),0)	This formula computes the lower control limit. If metric type is 1, then it is a general percent type of control, and the lower control limit is 2-sigma percent less than the center line. The 2-sigma value is Cell O2 in Figure 11 - Standard Deviations and Deviation from General Center Line. If the metric type is 3, then it is a by day of the week type control, and the lower control limit is 2-sigma percent less than the modified center line. The 2-sigma value and the modified center line value are looked up from the table in Figure 11 - Standard Deviations and Deviation from General Center Line. If the metric is type 2, then it is a scattered type control. The lower control limit is cell O2 in Figure 11 - Standard Deviations and Deviation from General Center Line. All lower control limits are limited not to go beyond zero.

Column	Formula	Comment

H	=MIN(MAX(IF(ISNUMBER(A2),
	IF(MetricType=1,D2+D2*HLOOKUP("Sun",MetDOW,2,FALSE),
	IF(MetricType=2,HLOOKUP("Mon",MetDOW,2,FALSE),
	IF(MetricType=3,F2+F2*HLOOKUP(E2,MetDOW,2,FALSE),
	NA()))),NA()),0),MaxVal)
		This formula computes the lower control limit.
		
		If metric type is 1, then it is a general percent type of control, and the lower control limit is 2-sigma percent more than the center line.
		
		The 2-sigma value is Cell O2 in Figure 11 - Standard Deviations and Deviation from General Center Line.
		
		If the metric type is 3, then it is a by day of the week type control, and the lower control limit is 2-sigma percent more than the modified center line.
		
		The 2-sigma value and the modified center line value are looked up from the table in Figure 11 - Standard Deviations and Deviation from General Center Line.
		
		If the metric is type 2, then it is a scattered type control. The upper control limit is cell P2 in Figure 11 - Standard Deviations and Deviation from General Center Line.
		
		Upper control limits are not allowed to exceed 9.99E+17 for numerical data or 100% for percent data.

Conditional formatting is applied to highlight the cell in green if the value is below the lower control limit and in red if the value is above the upper control limit. This color scheme is carried over to the chart where:

- Raw data is plotted as a dark blue line with diamonds indicating the individually measured points.
- The center line is plotted as a solid yellow line.
- The lower control limit is plotted as a solid green line.
- The upper control limit is plotted as a solid red line.

The two other type controls and sample computations are shown below.

Figure 13 - By Day of the Week Control Chart

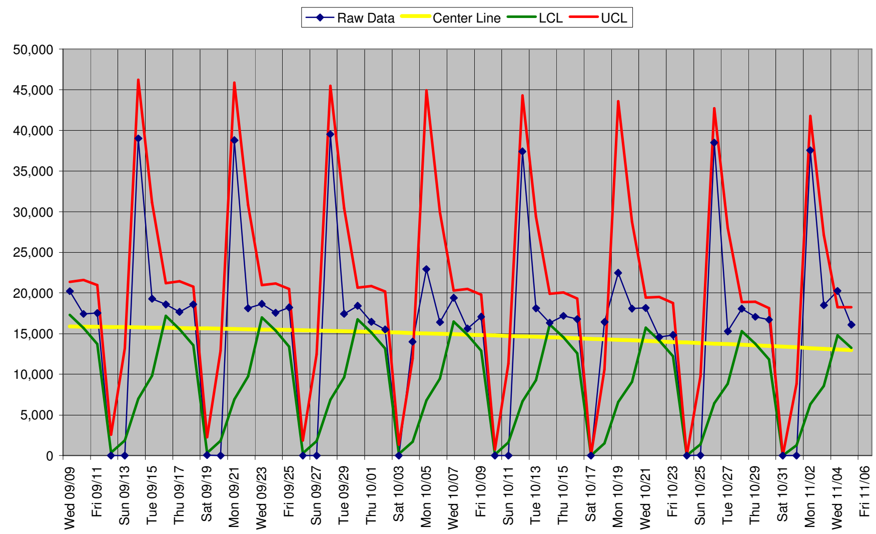


Figure 14 - Sample Computations for Day of the Week Control Chart

Date	Value	E-Date	Center Line	DOW	New CL	LCL	UCL
09/09/09	20,223	41	15,878	Wed	19,345	17,319	21,371
09/10/09	17,426	42	15,859	Thu	18,654	15,690	21,618
09/11/09	17,533	43	15,839	Fri	17,347	13,701	20,994
09/12/09	0	44	15,817	Sat	1,474	368	2,579
09/13/09	0	45	15,795	Sun	7,547	1,887	13,207
09/14/09	39,011	46	15,772	Mon	26,595	6,970	46,219
09/15/09	19,288	47	15,748	Tue	20,477	9,823	31,131
09/16/09	18,596	48	15,722	Wed	19,189	17,179	21,199

Figure 15 - Scattered Control Chart

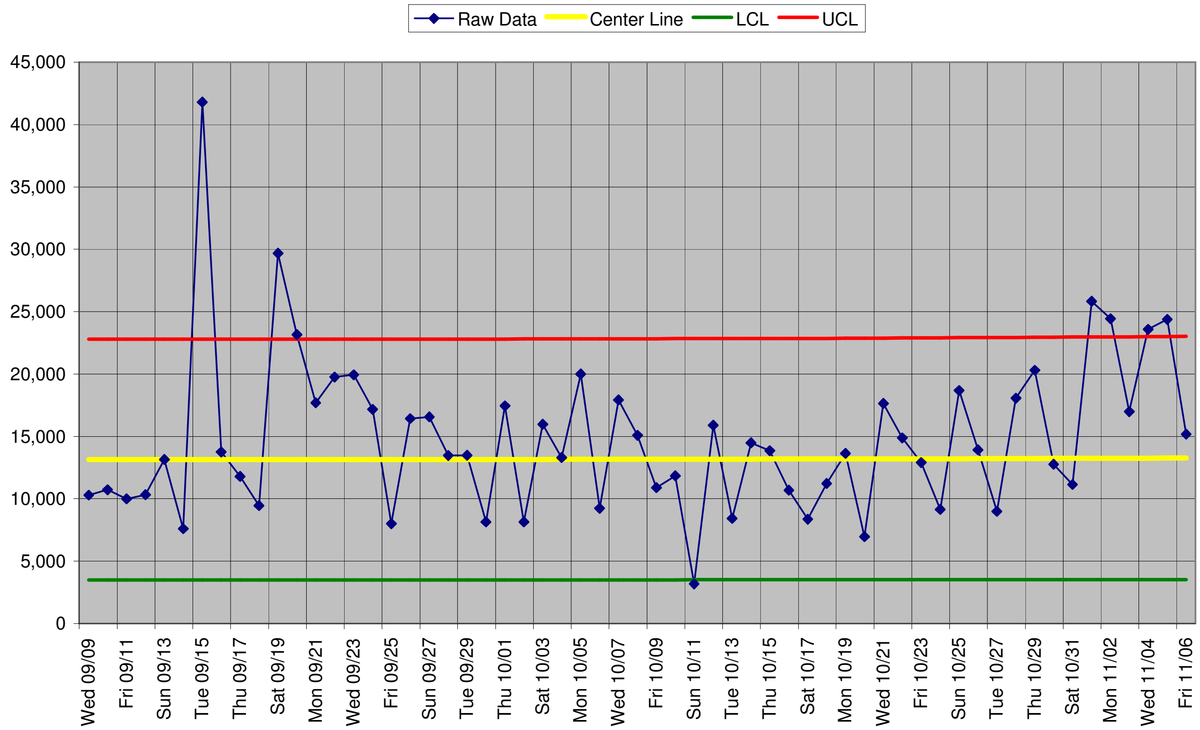


Figure 16 - Sample Computations for Scattered Control Chart

	A	B	C	D	E	F	G	H
1	Date	Value	E-Date	Center Line	DOW	New CL	LCL	UCL
2	09/09/09	10,289	39	13,147	Wed	13,147	3,493	22,801
3	09/10/09	10,706	40	13,147	Thu	13,147	3,493	22,800
4	09/11/09	9,989	41	13,147	Fri	13,147	3,493	22,800
5	09/12/09	10,328	42	13,147	Sat	13,147	3,493	22,800
6	09/13/09	13,149	43	13,146	Sun	13,146	3,493	22,800
7	09/14/09	7,586	44	13,146	Mon	13,146	3,493	22,800
8	09/15/09	41,794	45	13,146	Tue	13,146	3,493	22,799

Appendix A Least Squares Fit of a Third Degree Polynomial

This VB code can be used in conjunction with Excel to produce the coefficients for a least squares fit third degree polynomial based on the observed data. The method uses Cramer's Rule to solve the simultaneous equations. See notes at the end for limitations.

Coefficients can also be calculated using the LINEST() function in Excel, but this would require adding multiple additional columns to the spreadsheet.

```
Option Base 1

Public Function cubest(MyX As Range, MyY As Range) As Variant
Application.Volatile
Dim NumRows As Long
Dim k As Long
Dim SigY As Double, SigX As Double, SigX2 As Double, SigX3 As Double, SigX4 As Double, _
    SigX5 As Double, SigX6 As Double, SigXY As Double, SigX2Y As Double, SigX3Y As Double, N As
Double
Dim DetA As Double
Dim Result As Variant
Dim AnsA As Variant, AnsB As Variant, AnsC As Variant, AnsD As Variant

'Check data

If MyX.Rows.Count <> MyY.Rows.Count Then
    AnsA = "Row Mismatch"
    AnsB = "Number X rows = " & MyX.Rows.Count
    AnsC = "Number Y rows = " & MyY.Rows.Count
    AnsD = "***"
    GoTo MyEnd
End If

' initialize variables
SigY = 0
SigX = 0
SigX2 = 0
SigX3 = 0
SigX4 = 0
SigX5 = 0
SigX6 = 0
SigXY = 0
SigX2Y = 0
SigX3Y = 0
N = MyX.Rows.Count


' calculate sigmas
For k = 1 To MyX.Rows.Count
    If Application.IsNumber(MyY.Cells(k, 1)) And Application.IsNumber(MyX.Cells(k, 1)) Then
        SigY = SigY + MyY.Cells(k, 1)
        SigX = SigX + MyX.Cells(k, 1)
        SigX2 = SigX2 + MyX.Cells(k, 1) ^ 2
```

```
        SigX3 = SigX3 + MyX.Cells(k, 1) ^ 3
        SigX4 = SigX4 + MyX.Cells(k, 1) ^ 4
        SigX5 = SigX5 + MyX.Cells(k, 1) ^ 5
        SigX6 = SigX6 + MyX.Cells(k, 1) ^ 6
        SigXY = SigXY + MyX.Cells(k, 1) * MyY.Cells(k, 1)
        SigX2Y = SigX2Y + MyX.Cells(k, 1) ^ 2 * MyY.Cells(k, 1)
        SigX3Y = SigX3Y + MyX.Cells(k, 1) ^ 3 * MyY.Cells(k, 1)
    End If
Next

'Compute Results

DetA = Det4(N, SigX, SigX2, SigX3, SigX, SigX2, SigX3, SigX4, SigX2, SigX3, SigX4, SigX5, SigX3, SigX4,
SigX5, SigX6)

If DetA = 0 Then
    AnsA = "Divide by Zero Error"
    AnsB = "Divide by Zero Error"
    AnsC = "Divide by Zero Error"
    AnsD = "Divide by Zero Error"
    GoTo MyEnd
End If

AnsD = Det4(SigY, SigXY, SigX2Y, SigX3Y, SigX, SigX2, SigX3, SigX4, SigX2, SigX3, SigX4, SigX5, SigX3,
SigX4, SigX5, SigX6) / DetA
AnsC = Det4(N, SigY, SigX2, SigX3, SigX, SigXY, SigX3, SigX4, SigX2, SigX2Y, SigX4, SigX5, SigX3,
SigX3Y, SigX5, SigX6) / DetA
AnsB = Det4(N, SigX, SigX2, SigX3, SigX, SigX2, SigX3, SigX4, SigY, SigXY, SigX2Y, SigX3Y, SigX3, SigX4,
SigX5, SigX6) / DetA
AnsA = Det4(N, SigX, SigX2, SigX3, SigX, SigX2, SigX3, SigX4, SigX2, SigX3, SigX4, SigX5, SigY, SigXY,
SigX2Y, SigX3Y) / DetA



MyEnd:
cubest = Application.Transpose(Array(AnsA, AnsB, AnsC, AnsD))

End Function

Function Det4(A11 As Double, A12 As Double, A13 As Double, A14 As Double, _
    A21 As Double, A22 As Double, A23 As Double, A24 As Double, _
    A31 As Double, A32 As Double, A33 As Double, A34 As Double, _
    A41 As Double, A42 As Double, A43 As Double, A44 As Double) As Double
Application.Volatile
Det4 = A11 * A22 * A33 * A44 - A11 * A22 * A34 * A43 + A11 * A23 * A34 * A42 - A11 * A23 * A32 *
A44 + A11 * A24 * A32 * A43 - A11 * A24 * A33 * A42 _
      - A12 * A23 * A34 * A41 + A12 * A23 * A31 * A44 - A12 * A24 * A31 * A43 + A12 * A24 * A33 *
A41 - A12 * A21 * A33 * A44 + A12 * A21 * A34 * A43 _
      + A13 * A24 * A31 * A42 - A13 * A24 * A32 * A41 + A13 * A21 * A32 * A44 - A13 * A21 * A34 *
A42 + A13 * A22 * A34 * A41 - A13 * A22 * A31 * A44 _
      - A14 * A21 * A32 * A43 + A14 * A21 * A33 * A42 - A14 * A22 * A33 * A41 + A14 * A22 * A31 *
A43 - A14 * A23 * A31 * A42 + A14 * A23 * A32 * A41
```

End Function

<u>NOTES</u>: Although the Elapsed Date (E-Date) is calculated to prevent overflow, there are certain conditions in which these overflows occur using this method. These conditions occur when:

- There are few values in the data set. For example, a change in process requires an evaluation of the last 30 days instead of the last 180 days.
- The E-Date values are relatively high.

The combination of less than 60 data points and E-dates in excess of 200 seem to trigger the overflow. Visually, the overflow manifests itself as the right shaped curve but in the wrong position (too high or too low).

The work around for this condition is to use smaller E-dates. Instead of setting the Baseline date to 1/1/2009, change the Baseline date to the formula: =TODAY()-90. When getting the last 180 days of data the E-Date value will range from -90 to +90 and the overflow condition apparently goes away.