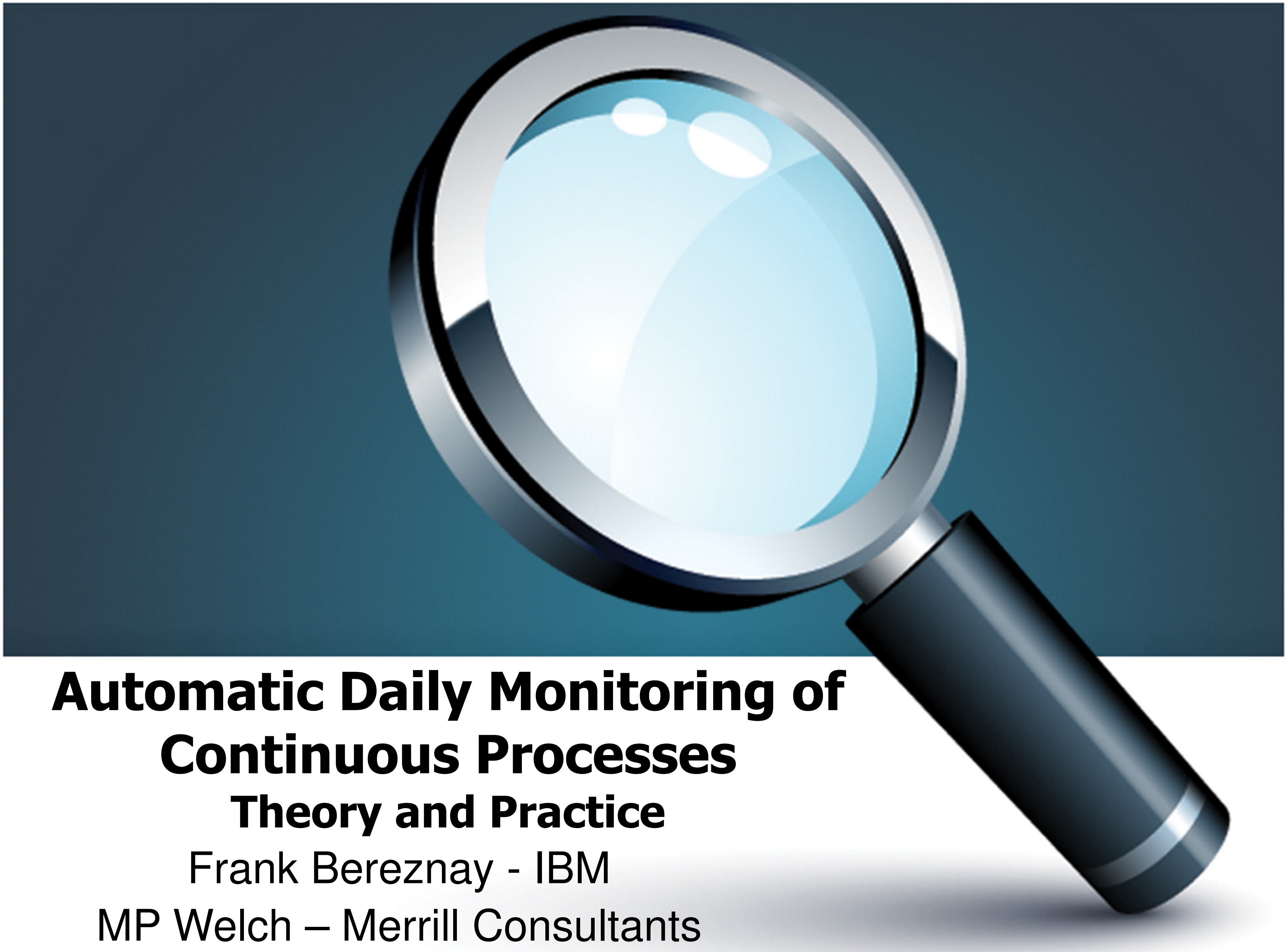


Automatic Daily Monitoring of Continuous Processes

Theory and Practice

Frank Bereznay - IBM

MP Welch – Merrill Consultants

Agenda

- Theory
  - Populations and Samples
  - Classical Statistical Inference and Control Charts
  - Our World
  - Multivariate Adaptive Statistical Filtering (MASF)
    - Reference Sets
    - Aggregation Policies / Clustering Techniques
  - Users of MASF
    - Trubin
  - Other Methods
    - Kaminsky – Pathology Detection
    - Seliverstov – Stock Market Trending Techniques

Populations and Samples

- A Phrase worth remembering!
  - Populations have Parameters.
  - Samples have Statistics.
- A population is a set of data points with fixed boundaries.
- A sample is a <u>random</u> subset of a population.
- Classical statistical inference deals with estimating population parameters by taking samples and calculating statistics.

Wine Bottling Example

- Bottling machine fills 750 ml wine bottles.
  - Machine is engineered such that 99.7% of the time the amount of wine dispensed should be 750 +/- 1 ml.
- A bottling run of 5 barrels is planned.
  - 1,500 wine bottles.
  - Approximately 2 pallets worth of wine.

Wine Trivia

75	grapes	=	1 cluster
1	cluster	=	1 glass
4	glasses	=	1 bottle
12	bottles	=	1 case
300	bottles	=	1 barrel

Wine Bottling Example

- What is our data population?

Wine Bottling Example

- What is our data population?
  - 1,500 bottles.

Wine Bottling Example

- What is our data population?
  - 1,500 bottles
- What are it’s parameters?

Wine Bottling Example

- What is our data population?
  - 1,500 bottles
- What are it’s parameters?
  - Mean is 750 ml.
  - Standard Deviation is .33 ml.

Wine Bottling Example

- What is our data population?
  - 1,500 bottles
- What are it's parameters?
  - Mean is 750 ml
  - Standard Deviation is .33 ml
- How can we validate the bottling machine is working properly?
  - We want to be 99% certain the machine is working properly.
  - Manually measure the volume of 7 randomly selected bottles.
    - Sample size a function of Standard Deviation and error measure.

      $N = (z_{\alpha/2})^2\sigma^2/E^2$
  - Conduct a two tailed Hypothesis test about the population mean.

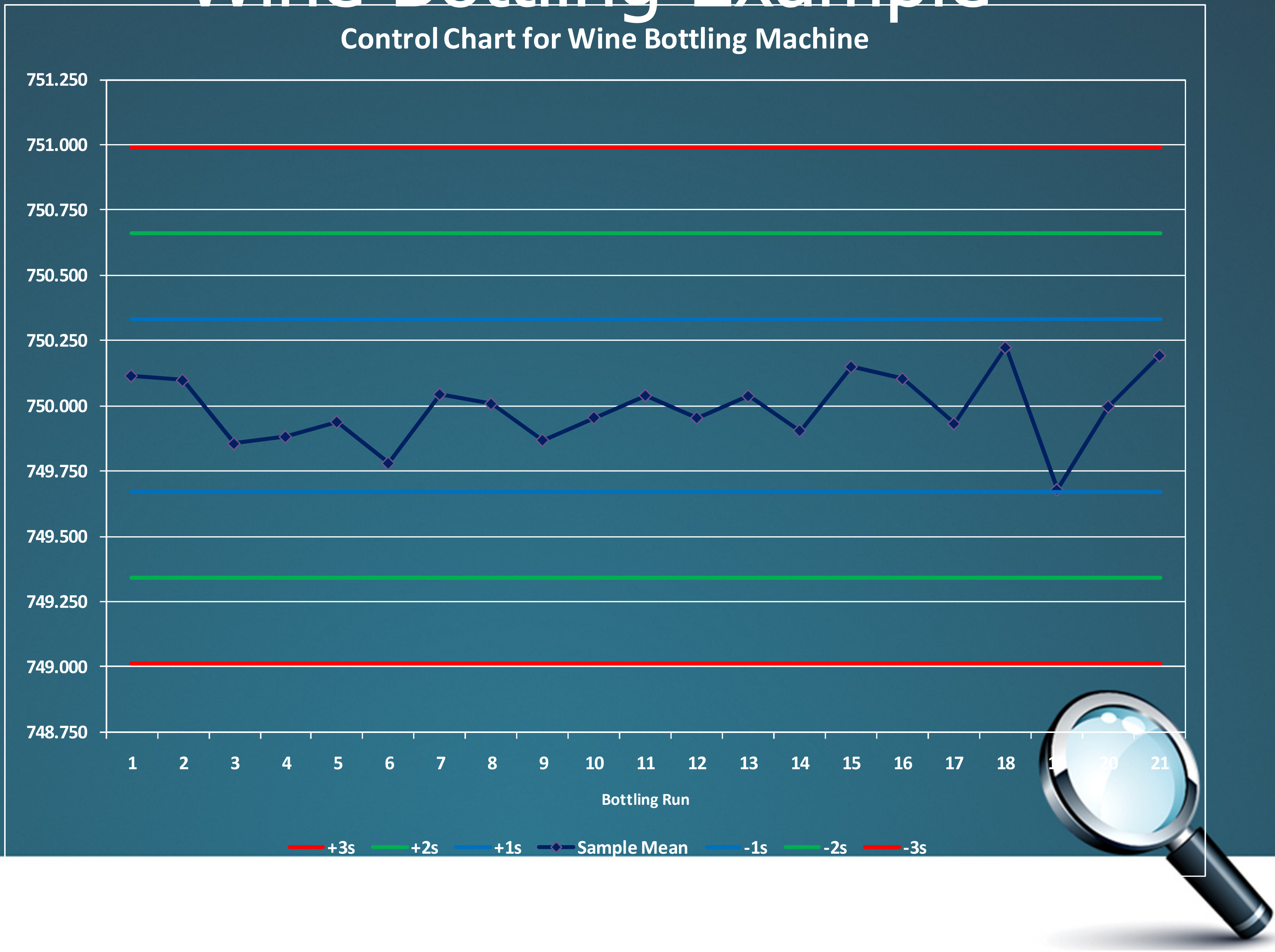
Wine Bottling Example

Bottle #	Ml of Wine		Population Parameters					
1	749.901		Average	749.996				
2	749.578		Std Dev	0.335				
3	750.081					Sample	Bottle #	
4	750.421					1	1245	749.356
5	750.395					2	1287	750.021
6	750.572					3	1500	750.271
7	749.279					4	128	750.176
8	749.923					5	82	750.476
9	750.361					6	17	749.813
10	749.641					7	866	749.989
11	749.772							
12	749.442					Sample Statistics		
13	749.391						Average	750.015
14	749.677						Std Dev	0.361
15	749.745							
16	749.301						t-Value	0.116
17	749.813						95%	1.960
18	749.867						99%	2.576
1495	749.753							
1499	749.784							
1500	750.271							

Wine Bottling Example

- Conclusion from test
  - There is insufficient evidence to reject the null hypothesis that the mean of the data population is 750 ml.
  - This conclusion does not imply or mean that the population mean is 750 ml, it just states the evidence is not strong enough to reject it.
    - Nature of test.
    - Wesel wording.

Wine Bottling Example



Classical Statistical Inference

- Key Points
  - Based on repeatable events.
    - The expected value is the same for each event.
  - Well defined batches or data populations.
  - Samples can be randomly taken from a data population.

Not exactly the kind of world we work in!

Our World

- Large volumes of data.
- Lots of variability in the data.
- Typically time series data sets.
- Populations boundaries not always clear.
- Difficult to randomly sample time series data sets.

- Conclusion
  – Classical statistical inference not well suited to analyze IT infrastructure instrumentation metrics.

MASF

- MASF - Multivariate Adaptive Statistical Filtering.
- Seminal CMG paper written in 1995 by Buzen & Shum.
- Directly addresses the challenges and problems faced by Capacity Management professionals when working with operational metrics.
- Provides a different analytical framework to work with metrics.
- Suggests reporting formats for working with large groups of resources.

A must read paper for all Capacity Management practitioners!

Available on the www.cmg.org website.

MASF

- Compare and contrast with sampling theory.
  - A fixed data population is replaced with a body of data that moves and changes over time.
  - A single random sample is replaced by multiple reference sets.
    - Random sample – Best possible data to estimate population parameters.
    - Reference set – A period of typical operation.
- Reference sets are aggregated across periods of typical operation to become an Adaptive Filtering Policy.
  - Between 10 to 20 reference set data points are needed.
    - You want the data to reflect recent experience.

Reference Set

- Typically each data point is one hour duration.
- A week can contain as many as 168 (24x7) reference sets.
  - Each hour of each day will be a separate measurement point.
  - A week provides one data point for each of the 168 reference sets.

Day/Hour	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023	024
Tue	025	026	027	028	029	030	031	032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048
Wed	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064	065	066	067	068	069	070	071	072
Thur	073	074	075	076	077	078	079	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095	096
Fri	097	098	099	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Sat	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
Sun	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168

Combining Reference Sets Across Weeks

- A key part of the MASF methodology.
- This implies that the workload is longitudinally stable.
- A conceptual design assumption that should be verified.

Mainframe LPAR Reference Set

Ref_Set	Hour	Data Points	CPU Mean	CPU Std
200	0	17	20.3	3.4
201	1	17	32.0	4.3
202	2	17	17.1	1.7
203	3	17	8.6	1.4
204	4	17	7.9	0.9
205	5	17	10.4	0.7
206	6	17	14.3	1.1
207	7	17	23.7	2.9
208	8	17	43.9	7.3
209	9	17	58.0	10.4
210	10	17	57.4	10.6
211	11	17	53.2	10.2
212	12	17	40.2	7.0
213	13	17	42.0	7.4
214	14	17	48.3	10.4
215	15	17	47.2	10.4
216	16	17	40.2	8.3
217	17	17	36.7	4.1
218	18	17	29.9	2.7
219	19	17	14.7	1.5
220	20	17	15.3	1.5
221	21	17	35.3	9.6
222	22	17	47.7	7.4
223	23	17	30.6	5.8

Mainframe LPAR Reference Set

Ref_Set	Hour	Data Points	CPU Mean	CPU Std
200	0	17	20.3	3.4
201	1	17	32.0	4.3
202	2	17	17.1	1.7
203	3	17	8.6	1.4
204	4	17	7.9	0.9
205	5	17	10.4	0.7
206	6	17	14.3	1.1
207	7	17	23.7	2.9
208	8	17	43.9	7.3
209	9	17	58.0	10.4
210	10	17	57.4	10.6
211	11	17	53.2	10.2
212	12	17	40.2	7.0
213	13	17	42.0	7.4
214	14	17	48.3	10.4
215	15	17	47.2	10.4
216	16	17	40.2	8.3
217	17	17	36.7	4.1
218	18	17	29.9	2.7
219	19	17	14.7	1.5
220	20	17	15.3	1.5
221	21	17	35.3	9.6
222	22	17	47.7	7.4
223	23	17	30.6	5.8

Ref_Set	Hour	Data Points	CPU Mean	CPU Std
200	0	10	19.6	1.6
201	1	10	31.9	4.6
202	2	10	18.0	1.0
203	3	10	8.1	0.6
204	4	10	8.1	0.8
205	5	10	10.4	0.7
206	6	10	14.2	1.2
207	7	10	23.8	3.1
208	8	10	44.2	7.6
209	9	10	58.6	11.3
210	10	10	57.6	11.4
211	11	10	53.9	10.8
212	12	10	40.4	7.6
213	13	10	42.4	7.6
214	14	10	48.7	11.0
215	15	10	47.6	10.6
216	16	10	40.4	8.8
217	17	10	36.8	4.3
218	18	10	30.0	3.4
219	19	10	15.1	1.6
220	20	10	15.8	1.5
221	21	10	40.2	6.5
222	22	10	43.1	1.9
223	23	10	29.8	4.0

Validating Longitudinal Stability

- Previous example was a simple visual inspection of the two tables.
  - The artist approach.
- A better approach would be to measure the differences against established thresholds.
  - Consider the workload stable if the mean difference is less than 1.5% and the variance difference is less than 2.5% for 90% of the Reference Sets of interest.
  - The scientist approach.

Validating Longitudinal Stability

Ref_Set	Date	Hour	Data Points	CPU Mean	CPU Std		Data Points	CPU Mean	CPU Std		Delta Mean	Delta Std
200	1/3/2011	0	17	20.3	3.4		10	19.6	1.6		0.7	1.8
201	1/3/2011	1	17	32.0	4.3		10	31.9	4.6		0.1	-0.3
202	1/3/2011	2	17	17.1	1.7		10	18.0	1.0		-1.0	0.8
203	1/3/2011	3	17	8.6	1.4		10	8.1	0.6		0.5	0.8
204	1/3/2011	4	17	7.9	0.9		10	8.1	0.8		-0.1	0.1
205	1/3/2011	5	17	10.4	0.7		10	10.4	0.7		0.0	0.0
206	1/3/2011	6	17	14.3	1.1		10	14.2	1.2		0.1	-0.1
207	1/3/2011	7	17	23.7	2.9		10	23.8	3.1		-0.1	-0.2
208	1/3/2011	8	17	43.9	7.3		10	44.2	7.6		-0.3	-0.4
209	1/3/2011	9	17	58.0	10.4		10	58.6	11.3		-0.6	-0.9
210	1/3/2011	10	17	57.4	10.6		10	57.6	11.4		-0.2	-0.8
211	1/3/2011	11	17	53.2	10.2		10	53.9	10.8		-0.7	-0.6
212	1/3/2011	12	17	40.2	7.0		10	40.4	7.6		-0.2	-0.7
213	1/3/2011	13	17	42.0	7.4		10	42.4	7.6		-0.4	-0.3
214	1/3/2011	14	17	48.3	10.4		10	48.7	11.0		-0.4	-0.6
215	1/3/2011	15	17	47.2	10.4		10	47.6	10.6		-0.4	-0.2
216	1/3/2011	16	17	40.2	8.3		10	40.4	8.8		-0.3	-0.4
217	1/3/2011	17	17	36.7	4.1		10	36.8	4.3		-0.1	-0.1
218	1/3/2011	18	17	29.9	2.7		10	30.0	3.4		-0.1	-0.7
219	1/3/2011	19	17	14.7	1.5		10	15.1	1.6		-0.4	-0.2
220	1/3/2011	20	17	15.3	1.5		10	15.8	1.5		-0.4	-0.1
221	1/3/2011	21	17	35.3	9.6		10	40.2	6.5		-5.0	3.1
222	1/3/2011	22	17	47.7	7.4		10	43.1	1.9		4.5	5.6
223	1/3/2011	23	17	30.6	5.8		10	29.8	4.0		0.8	1.8

Validating Longitudinal Stability

- Conclusion.
  - Using the 1.5%/2.5% criteria, this is a very stable workload week over week. Only two non prime Reference Sets exceeded the threshold.
- Other calculation schemes are possible.
- There is no single right or wrong answer here.
- Longitudinal Stability is another way of saying this is a period of typical operation.
- The MASF framework also recognizes the need for seasonal adjustments to resource consumption profiles.
- Ultimately you will need to make a judgment call about the stability of the workload over time.

Adaptive Filtering Policy

- Building an Adaptive Filtering Policy involves aggregating each reference set across multiple weeks of typical operation.
  - Reference Set #001 – 0000 Monday
    - Data Point #01 – 02 May 2011
    - Data Point #02 – 09 May 2011
    - Data Point #03 – 16 May 2011
    - Repeat 17 times
  - Reference Set #002 - #168
    - Repeat above steps
- This process will take 20 weeks to develop the policy.
  - Too long a period for most workloads, not good.
  - Need to aggregate similar Reference Sets within a week.

Combining Reference Sets within a Week

- Computationally similar to validating longitudinal stability.
  - Look for Reference Sets with similar mean and variance values for a given week.
  - The consolidations should be done within windows of common usage patterns.
    - Example – Don’t consolidate a prime shift online workload with an off prime shift batch workload even if they have similar resource consumptions profiles.

Combining Similar Reference Sets

- One possible computational framework.
  - Divide the week into the following groups.
    - Four 6 hour shifts – 00-05,06-11,12-17,18-23
    - Weekday and Weekend groups.
    - A total of 8 groups
- Within each group look for consolidation opportunities.

Day/Hour	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Tue	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323
Wed	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423
Thur	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523
Fri	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623
Sat	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723
Sun	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123

Combining Similar Reference Sets

- The table now contains the mean value for each Reference Set.

Day/Hour	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	19.6	31.9	18.0	8.1	8.1	10.4	14.2	23.8	44.2	58.6	57.6	53.9	40.4	42.4	48.7	47.6	40.4	36.8	30.0	15.1	15.8	40.2	43.1	29.8
Tue	38.1	38.4	26.7	14.4	9.4	11.1	14.3	23.1	43.4	57.3	57.9	53.1	39.6	43.6	50.0	48.5	41.2	38.7	30.6	15.0	15.6	41.6	43.4	29.7
Wed	42.8	38.9	27.7	12.9	8.8	10.6	13.4	21.2	41.5	54.2	54.7	50.6	37.2	41.5	47.6	46.2	40.1	36.6	28.4	15.0	15.7	35.2	43.8	27.2
Thur	45.5	45.1	28.4	13.7	9.4	10.4	13.3	21.5	40.3	52.9	52.9	48.4	36.1	40.2	45.5	44.1	37.9	23.7	16.6	13.8	14.5	40.9	42.9	28.9
Fri	44.3	41.0	29.6	12.8	8.7	11.3	13.3	20.2	37.3	48.6	49.9	45.9	34.3	38.8	43.1	41.2	34.7	20.3	14.8	13.1	13.6	40.2	39.1	27.9
Sat	38.1	38.0	28.6	17.3	11.5	8.9	9.3	11.4	14.1	14.4	14.6	13.3	12.0	15.2	13.4	12.4	10.9	18.1	19.4	11.8	10.7	39.3	29.7	15.8
Sun	19.7	13.5	11.1	13.7	8.9	14.7	16.2	17.2	19.5	16.7	17.5	13.7	12.1	12.9	10.5	11.4	10.7	21.5	22.6	14.4	26.5	26.6	33.1	27.2

Combining Similar Reference Sets

- Within each group divide the Reference Sets into a small and large variance sub group and order by mean value for each sub group.
- Start a consolidation group with the lowest value observation and add 1.5% to it's mean value. This becomes the upper limit for the current consolidation group.
- Iterate through the observations until the mean value of the nth observation exceeds the upper limit. All observations up to that point are in the same consolidation group. Start a new consolidation group with the observation that exceeded the upper limit.
- Repeat process for each variance sub group.

Combining Similar Reference Sets

Group	Ref Set	CPU Mean	CPU Std	Group	Ref Set	CPU Mean	CPU Std	Group	Ref Set	CPU Mean	CPU Std	Group	Ref Set	CPU Mean	CPU Std
1	204	8.1	0.8	2	506	13.3	1.4	3	617	20.3	0.6	4	619	13.1	0.7
1	203	8.1	0.6	2	606	13.3	0.9	3	517	23.7	1.2	4	620	13.6	0.9
1	604	8.7	2.0	2	406	13.4	0.7	3	612	34.3	4.5	4	519	13.8	1.2
1	404	8.8	1.2	2	206	14.2	1.2	3	616	34.7	1.0	4	520	14.5	0.8
1	504	9.4	1.1	2	306	14.3	1.6	3	512	36.1	1.4	4	618	14.8	0.9
1	304	9.4	1.5	2	607	20.2	1.2	3	417	36.6	1.3	4	419	15.0	0.9
1	205	10.4	0.7	2	407	21.2	1.1	3	217	36.8	4.3	4	319	15.0	0.9
1	505	10.4	1.0	2	507	21.5	1.2	3	412	37.2	1.1	4	219	15.1	1.6
1	405	10.6	1.2	2	307	23.1	1.5	3	516	37.9	1.0	4	320	15.6	0.6
1	305	11.1	1.5	2	207	23.8	3.1	3	317	38.7	1.3	4	420	15.7	1.5
1	605	11.3	1.6	2	608	37.3	1.3	3	613	38.8	4.2	4	220	15.8	1.5
1	603	12.8	1.8	2	508	40.3	1.3	3	312	39.6	1.5	4	518	16.6	0.6
1	403	12.9	1.9	2	408	41.5	1.7	3	416	40.1	1.1	4	423	27.2	4.1
1	503	13.7	0.8	2	308	43.4	1.7	3	513	40.2	1.1	4	623	27.9	0.5
1	303	14.4	4.7	2	208	44.2	7.6	3	216	40.4	8.8	4	418	28.4	1.9
1	202	18.0	1.0	2	611	45.9	1.7	3	212	40.4	7.6	4	523	28.9	1.4
1	200	19.6	1.6	2	511	48.4	1.1	3	615	41.2	1.6	4	323	29.7	1.3
1	302	26.7	6.5	2	609	48.6	1.7	3	316	41.2	0.9	4	223	29.8	4.0
1	402	27.7	4.3	2	610	49.9	1.5	3	413	41.5	1.3	4	218	30.0	3.4
1	502	28.4	2.2	2	411	50.6	2.2	3	213	42.4	7.6	4	318	30.6	2.3
1	602	29.6	4.0	2	509	52.9	1.8	3	614	43.1	3.2	4	421	35.2	2.0
1	201	31.9	4.6	2	510	52.9	1.2	3	313	43.6	1.8	4	622	39.1	0.9
1	300	38.1	7.9	2	311	53.1	1.8	3	515	44.1	1.3	4	621	40.2	1.4
1	301	38.4	4.6	2	211	53.9	10.8	3	514	45.5	1.0	4	221	40.2	6.5
1	401	38.9	5.2	2	409	54.2	1.3	3	415	46.2	2.1	4	521	40.9	0.9
1	601	41.0	4.9	2	410	54.7	1.2	3	215	47.6	10.6	4	321	41.6	0.8
1	400	42.8	7.3	2	309	57.3	2.2	3	414	47.6	1.6	4	522	42.9	1.4
1	600	44.3	6.0	2	310	57.9	1.7	3	315	48.5	2.1	4	222	43.1	1.9
1	501	45.1	2.5	2	210	57.6	11.4	3	214	48.7	11.0	4	322	43.4	1.4
1	500	45.5	4.7	2	209	58.6	11.3	3	314	50.0	1.9	4	422	43.8	2.3

Combining Similar Reference Sets

- The consolidated Reference Sets become Computational Groups.
- Calculate control limits for each Computational Group.
  - Simple mean and standard deviation typically used.
  - Other more sophisticated methods can also be used.
    - Favor recent events over older data.
- Map these values back to the Reference Sets they came from.

- We now have an Adaptive Filtering Policy!
  - Use this to evaluate future period activity.
  - May be adjusted to reflect seasonal activity.

Combining Similar Reference Sets

Compute Group	Data Points	Metric Mean	Metric Std
1	60	8.8	1.3
2	50	10.8	1.2
3	30	13.1	1.6
4	20	19.0	1.5
5	20	29.0	3.2
6	10	45.1	2.5
7	50	13.8	1.3
8	30	21.0	1.2
9	20	24.0	1.7
10	10	37.3	1.3
11	20	40.9	1.6
12	10	43.4	1.7
13	20	46.1	1.5
14	30	49.0	1.5
59	10	39.3	0.6
60	10	14.4	4.7
61	20	27.2	5.4
62	10	32.2	4.3
63	30	38.5	5.9
64	20	41.9	6.1
65	20	44.9	5.3
66	10	34.3	4.5
67	10	9.2	4.4
68	10	11.8	5.4
69	10	13.8	8.0

Sample Adaptive Filtering Policy

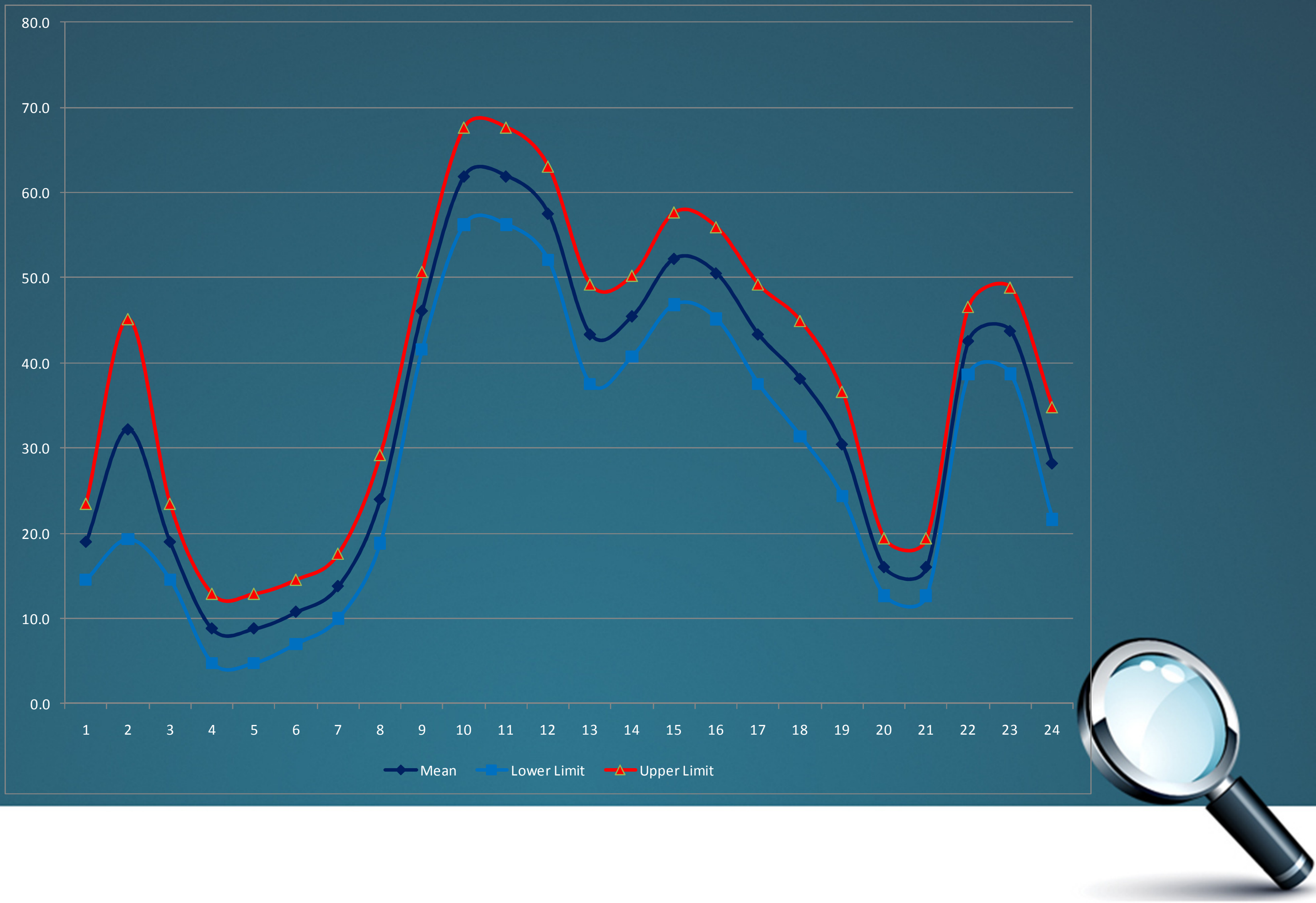
Reference Set	Metric Mean	Metric Std	Compute Group	Data Points
200	19.0	1.5	4	20
201	32.2	4.3	62	10
202	19.0	1.5	4	20
203	8.8	1.3	1	60
204	8.8	1.3	1	60
205	10.8	1.2	2	50
206	13.8	1.3	7	50
207	24.0	1.7	9	20
208	46.1	1.5	13	20
209	61.9	1.9	18	20
210	61.9	1.9	18	20
211	57.6	1.8	17	30
212	43.4	1.9	24	50
213	45.5	1.6	25	30
214	52.3	1.8	28	10
215	50.5	1.8	27	20
216	43.4	1.9	24	50
217	38.1	2.3	22	50
218	30.4	2.0	32	30
219	16.0	1.1	30	50
220	16.0	1.1	30	50
221	42.6	1.3	35	40
222	43.8	1.7	36	20
223	28.2	2.2	31	50

Sample Adaptive Filtering Policy

- Typical Use
  - Create an Adaptive Filtering Policy weekly.
    - Use 8 to 10 weeks of historical data.
    - Supplement with any seasonal or holiday adjustments.
  - Use the policy to evaluate one weeks worth of actual data.
  - Repeat process.

- There are many other ways this framework can be used!
- No right or wrong answer here.

Adaptive Filtering Policy



MASF Review

- Different conceptual framework for handling metrics.
- Divide and conquer the dynamic metrics we work with.
- Treat each hour of each day as a separate data population (aka Reference Set) and aggregate over time.
- Look for similar hours within shifts and combine Reference Sets to create Computational Groups.
- Use Computational Groups to create limits and map them back to the Reference Sets they represent.
- Apply any seasonal adjustment to the control limits.
- The result is an Adaptive Filtering Policy that will be used to evaluate future period activity.

The Master of MASF

- Igor Trubin – IBM
  - Exception Detection System, Based on the Statistical Process Control Concept – 2001
  - Global and Application Level Exception Detection System, Based on MASF Technique – 2002
  - Disk Subsystem Capacity Management, Based on Business Drivers, I/O Performance and MASF – 2003
  - Global and Application Level Exception Detection System, Based on MASF Technique – 2004
  - Capturing Workload Pathology by Statistical Exception Detection System – 2005

The Master of MASF

- System Management by Exception, Part 6 – 2006
- System Management by Exception: The Final Part – 2007

- Blog – itrubin.blogspot.com

Igor's Blog

SYSTEM MANAGEMENT BY EXCEPTION

HOW STATISTICAL FILTERING TECHNIQUES SUCH AS SPC, MASF, SIX SIGMA AND SEDS (STATISTICAL EXCEPTION DETECTION) ARE USED FOR CAPACITY MANAGEMENT ... BY IGOR TRUBIN

SEARCH THIS BLOG

Search

powered by Google™

BLOG ARCHIVE

- ► 2007 (2)
- ► 2008 (2)
- ► 2009 (15)
- ► 2010 (17)
- ▼ 2011 (4)

MONDAY, APRIL 25, 2011

UCL=LCL : How many standard deviations do we use for Control Charting? Use ZERO!

How many standard deviations do we use for upper (UCL) and lower (LCL) limits calculations on a control charts? 3? 1? **What about 0 st. dev.!?** Indeed, the simplest way to build MASF data for exception detection is to use 168 weekly hours averages as a baseline, so that would be the case when ZERO st. Dev is used to make UCL=LCL! Plus for further simplification the current data could be included in wider historical baseline (Why not?). My EV meta-metric in this case would be just difference between actual metric value and the average over baseline!

Related Papers

- Ron Kaminski – Kimberly Clark
  - Automating Process and Workload Pathology Detection
  - Automating Process Pathology Detection Rule Engine Design Hints
  - Time Stable Workload Characterization Techniques
  - Automating Workload Characterization by Policy
  - Business Metrics and Capacity Planning
- Dima Seliverstov – BMC Software
  - Application of Stock Market Technical Analysis Techniques for Computer System Performance Data
  - Applicability of Spectral Analysis Techniques to Computer Performance Data Analysis

Summary

- The primary objective here is to identify analytical techniques that will allow us to monitor large pools of servers and other infrastructure assets to look for discernable change to their usage patters.
- If this was easy, it would have been done a long time ago.

- It will not work for all workloads, all the time.

Summary

- A positive note
  - Even if workloads are not repeatable and can't be monitored by this technique, displaying the mean and control limits for a workload provides valuable insights into it's behavior.
  - Reporting the variability of a variable along with its value is recommended for most management reporting efforts.