C7

"Data, data everywhere…

…and not a thought to think!"

Creating Organizational Data "Sanity"

Davis Balestracci

HARMONY

6844 Buckingham Ct.
Woodbury, MN 55125-2401

Phone: (651)-739-5356

e-mail: davisbharmony@mediaone.net

Institute for Healthcare Improvement Annual Forum

December 12, 2001

Davis Balestracci has had a variety of experiences in his 20-year career. The first half was industrial, most significantly with 3M, where he was awarded two corporate quality awards and two process technology awards for his innovative teaching and uses of statistical methods.

His interests then evolved to utilizing the Deming philosophy in management and service contexts. For the past 10 years, he has functioned as a Deming statistical consultant for two major Minnesota Twin Cities-based multi-specialty health care clinics, one with 500 physicians and 20 locations.

Davis has a B.S. degree in chemical engineering, an M.S. degree in statistics, yet describes himself as a "right-brained" statistician (He has also done graduate work in orchestral and choral conducting!).

Davis is a regular speaker at the IHI National Forums and is also known nationally and internationally for his passionate, provocative, challenging, yet humorous and down-to-earth style. He is well aware of the daily realities of implementing statistical approaches to quality and cultural transformation—including the inherent frustrations of dealing with "*those darn humans!*"

This year, he made the decision to become an independent consultant. The name of his company, HARMONY, reflects his melding of left-brain (analytical) and right-brain (psychological) approaches to quality as well as "*the passion of Beethoven composing symphonies*" with which he approaches his work—motivational and transformational consulting, seminars, retreats, and coaching.

In 1995, he was a member of a faculty team sponsored by the Harvard Institute for International Development that taught health care quality improvement methods in the Middle East to 80 health care leaders from Egypt, Palestine, Jordan, Morocco, and Israel. He subsequently acted as a consultant to the Palestinian Improvement Project in 1997 and at that time also presented a one-day seminar at The Technion in Haifa, Israel.

In summer 1998, he published a controversial special edition newsletter that was sent to 11,000 people, "Data 'Sanity': Statistical Thinking Applied to Everyday Data," for the Statistics Division of the American Society for Quality.

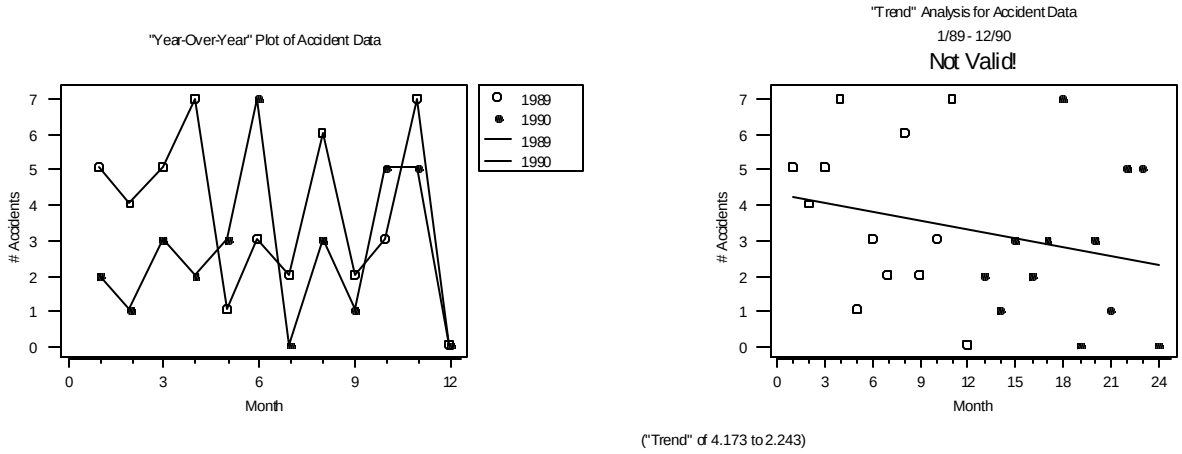
He is the author of a book summarizing his ideas, *Quality Improvement: Practical Applications for Medical Group Practice*, which is published by the Center for Research in Ambulatory Health Care Administration (CRAHCA) and currently in its 2nd edition.

Davis is a member of the American Society for Quality (ASQ), the Association for Quality and Participation (AQP), and is the past president of the Twin Cities Deming Forum. He can be reached by phone at (651)-739-5356 or via e-mail: davisbharmony@mediaone.net .

The "Safety Reward" Luncheon

You have been invited to attend a free pizza lunch in celebration of meeting a safety goal. Two years ago, your organization had 45 undesirable "incidents" and set a goal the past year of reducing them by at least 25%. The December data is in, and the yearly total was: 32 "incidents"—a 28.9% decrease! Also, look—Eight months were lower than the corresponding months in the previous year!

In fact, the "resident statistical expert" did a trend analysis and demonstrated that the decrease was more on the order of 46.2%! At this rate, 20 "incidents" are predicted for this year.



You think of all the hard work that takes place at the monthly safety meetings where each individual event is dissected and discussed to come up with a new safety policy. Then there are the months where you have zero accidents and the reasons for this are discussed and implemented.

Congratulations! All this hard work has paid off!

⇒ Well, actually, it hasn't…but, like they say, you can prove anything with statistics!

Well…that isn't necessarily true, either.

⇒ *Did you know that you can probably expect anywhere from 20-57 incidents in the next year if you continue this process with your hard-working people?!*

⇒ Did you also know that a different, *simpler*, analysis of this data contains enough information to *immediately* reduce the rate of these incidents by ***half***?

The QA Report Meeting [True Story]

Monthly Data			12-Month Period Summaries		
Arrests	**Vfib**	**Mo/Yr**	**Tot_Arr**	**Tot_Vfib**	**Period**
18	6	6/94	261	81	6/94-5/95
17	8	7/94	275	71	6/95-5/96
15	6	8/94			
19	6	9/94			
21	6	10/94			
21	8	11/94			
23	7	12/94			
25	7	1/95			
21	1	2/95			
30	9	3/95			
27	8	4/95			
24	9	5/95			
24	9	6/95			
19	2	7/95			
14	2	8/95			
21	7	9/95			
32	5	10/95			
19	4	11/95			
28	9	12/95			
28	10	1/96			
28	8	2/96			
17	5	3/96			
21	7	4/96			
24	3	5/96			
Note: Vfib is a term for ventricular fibrillation					

"We are running a slightly higher number of cardiac arrests per month. The total amount of cardiac arrests has risen from a mean of 21.75 (June 94- May 95), to 22.92 (June 95- May 96). This is an increase in 14 cardiac arrests in the last 12 months."

⇒ So…what are you going to do about it?!

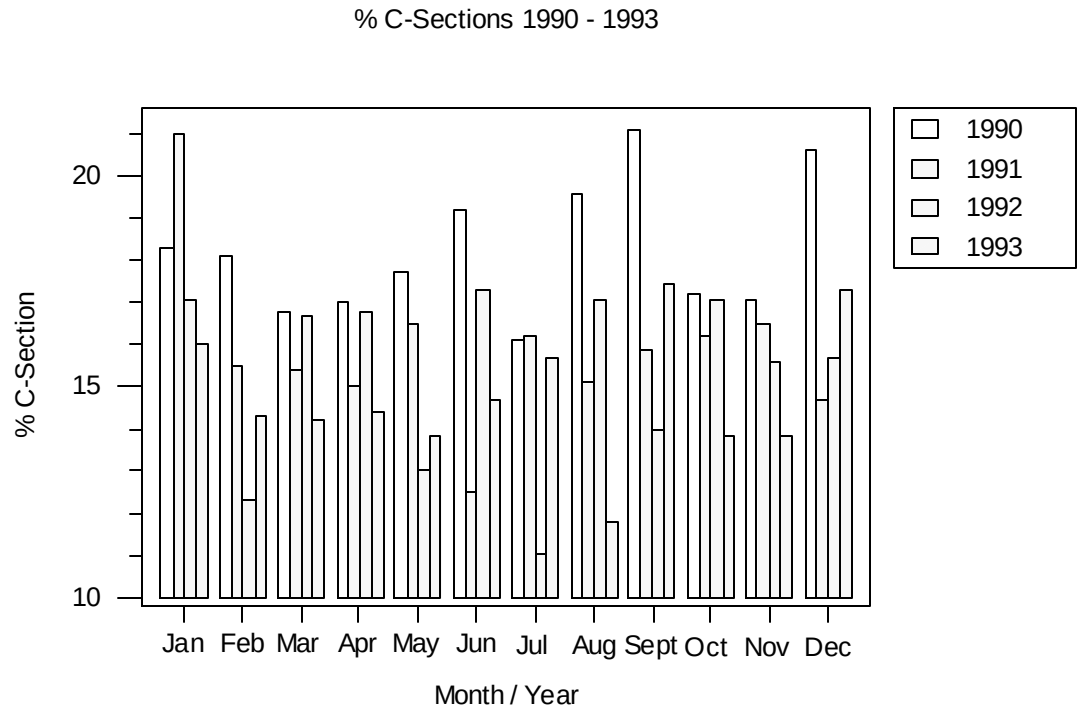
"Next we interpreted the data relating to Vfib Cardiac Arrests…This could be significant to our outcome, and…indicates a need for more sophisticated statistical analysis. It was already shown that the number of cardiac arrests has increased by a mean of 1.17 per month. Now we are adding to that increase, a decrease of times we are seeing Vfib as the initial rhythm. From June 1994 to May 1995 we arrived on scene to find Vfib as the initial rhythm with an overall mean of 6.75 times. That gave us a capture rate of 32.03%. This last year, June 1995 - May 1996, we are arriving to find Vfib as the initial rhythm with an overall mean of 5.92, and a capture rate of 25.81%. This obviously means that over the last year, we have responded to more cardiac arrests and found them in more advanced stages of arrest."

⇒ From this, one could conclude that since 275 is greater than 261, AND 71 is less than 81, there is some kind of crisis! Well…"Don't just stand there…DO something!"

What usually happens next?

"Oh, no…! The Monthly C-Section Meeting"

Your next meeting concerns a disturbing "trend" in C-sections…AGAIN!…(You'd rather be getting a cavity filled). JCAHO or NCQA also want you to establish a "standard." The all-too-familiar bar graph display is shown below.



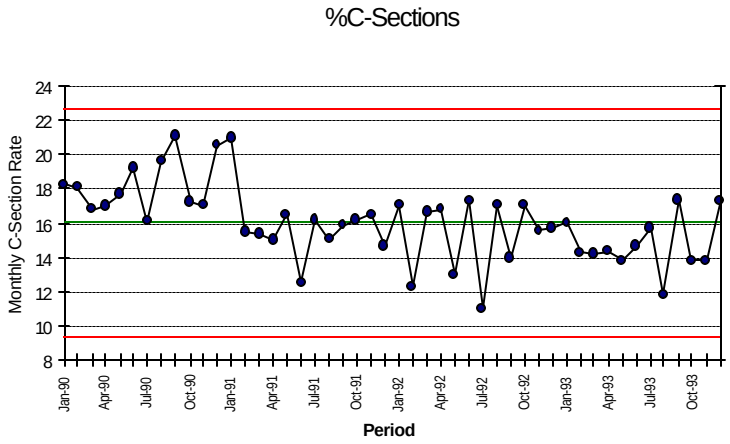
The "stats" are below, and, yes, the data are "normally distributed."

The "Stats"

	N	Mean	Median	TrMean	StDev	SE Mean	Min	Max	Q1	Q3
% C-Sect	48	16.046	16.150	16.030	2.230	0.322	11.000	21.100	14.475	17.175

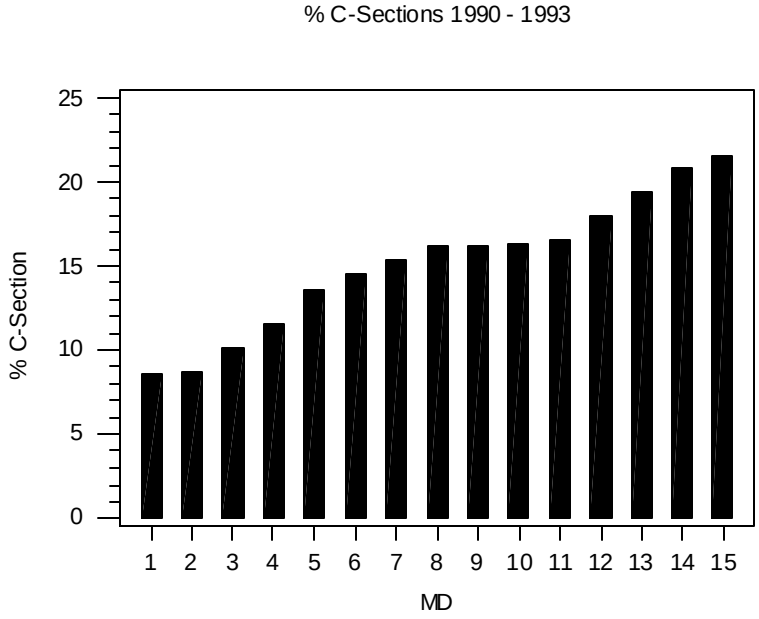
⇒ **So, what should the goal or standard be?**

An analyst thought the control chart below might be useful and concluded: "Since all the points are within the three sigma limits, there are no special causes. Nothing's really changed in four years."



This analyst also had some data on each of the 15 physicians who practice at the hospital:

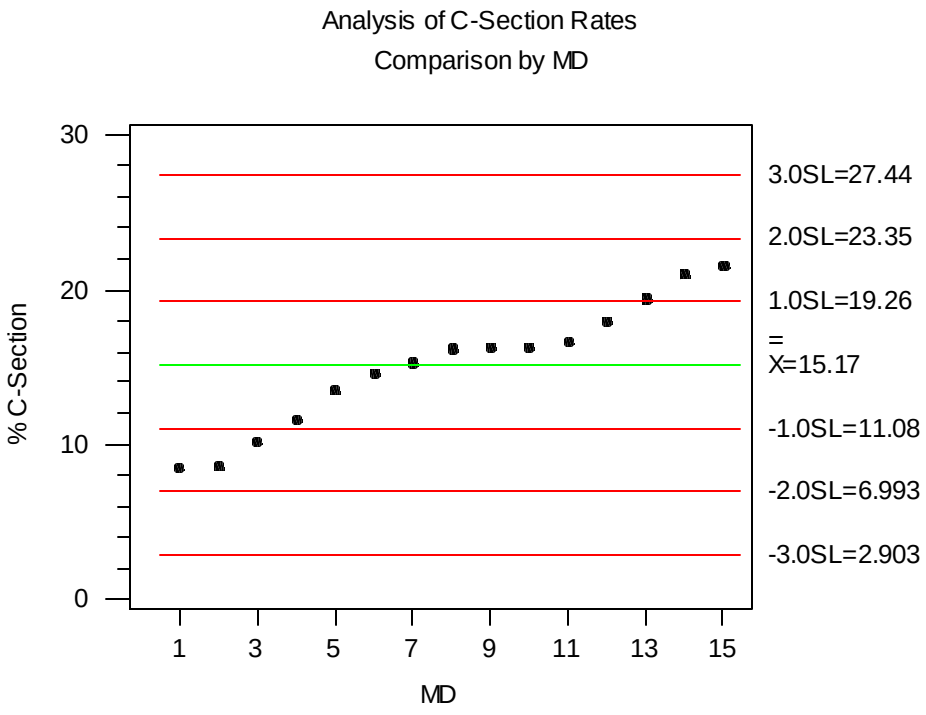
The individual rates are shown in ascending order in the bar graph below. How does this data factor in to the setting of the goal/standard? Should any of these physicians be investigated?



The "analyst" decides to do a more "statistical" analysis on this individual physician data. He adapted a previous protocol used by the pharmacy utilization committee (True story):

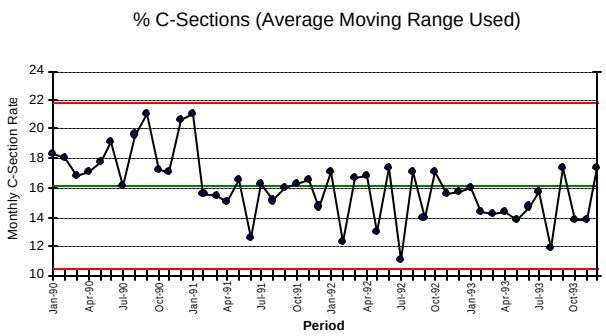
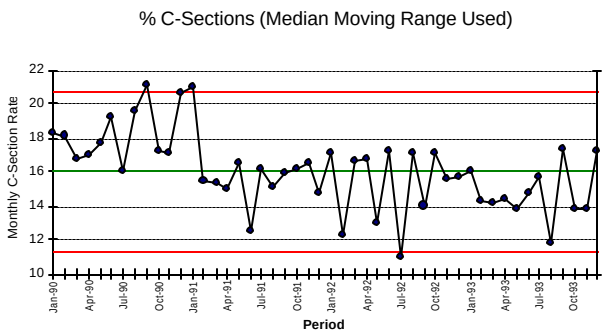
1. "Data will be tested for the normal distribution,"
2. "If the distribution is normal, physicians whose C-section rates deviate greater than one or two standard deviations from the mean are identified as outliers,"
3. "If the distributions is not normal, examine the distribution of data and establish an arbitrary cutoff point above which physicians should receive feedback (this cutoff point is subjective and variable based on the distribution of ratio data),"
4. "Anyone identified as an outlier will be reported to the medical director and given current literature on C-section research and published CQI projects from 'gold standard' benchmarked organizations."

The 15 physicians' data were tested for normality and "passed" (p-value = 0.584). This analysis is shown graphically below with "one," "two," and "three" standard deviation lines drawn in.



Another analyst noticed that the control chart limits on the first control chart were calculated using the overall standard deviation of the data (3x 2.23), which a "guru" at the IHI conference said must never, *never*, ***ever*** be done and recommended the **average moving range** to determine the limits.

A third analyst heard another IHI "guru" say to use the **median moving range**. From the charts below, it's obvious that this produced the most "accurate" chart because it exposes ***four*** opportunities that the other two analyses miss.

	Using the Average Moving Range seems to have tightened the limits by 1% on both sides. But…still no data points outside the limits.
	Using the Median Moving Range has tightened the limits 1% further than the Average Moving Range. All right!!…we see some points outside the three sigma limits! (How does this help the investigation?)

We are drowning in statistics! The answer must be in here somewhere! (*Well…actually, it isn't.*)

What usually happens now?

⇒ Did you know that all this discussion is moot?

What is proposed is "*simple…obvious…and WRONG!*"

So…how much time are you wasting in routine meetings because of such "**Data *IN*sanity?"**

❖ **Ask the Right Questions!**

"Data Inventory" Considerations

1. What is the **objective** of these data?
2. Is there an unambiguous **operational definition** to obtain a consistent numerical value for the process being measured?
   Is it **appropriate** for the stated objective?
3. How are these data **accumulated/collected**?
   Is the collection **appropriate** for the stated objective?
4. How are the data currently being **analyzed/displayed**?
   Is the analysis/display **appropriate**, given the way the data were collected?
5. What **action**, if any, is currently being taken with these data?

Given the objective and action, is anything "wrong" with the current number?

Practical

Accumulated

Records

Compilation

⇒Statistics: The ***art*** and science of ***collecting*** and analyzing data—asking the right questions.

⇒*Simple, efficient* data collection

⇒*Process*-oriented

"90% of statistics is half planning."

❖ The MBFC "Goal" Exercise:

An MBA has made a "business situation analysis" and come up with a formula to establish a critical organizational goal based on the number of employees:

If "N" is the number of employees:

1) The **monthly** goal is: $\frac{N}{4} + 0.433 \times \sqrt{N}$**, rounded *up* to the nearest "5" or "0"**

["Stretch" so as to attain the all-important quarterly goal],

2) The **quarterly** goal is: $3 \times \frac{N}{4} + 0.75 \times \sqrt{N}$**, rounded *up* as above** **["Crucial!"]**

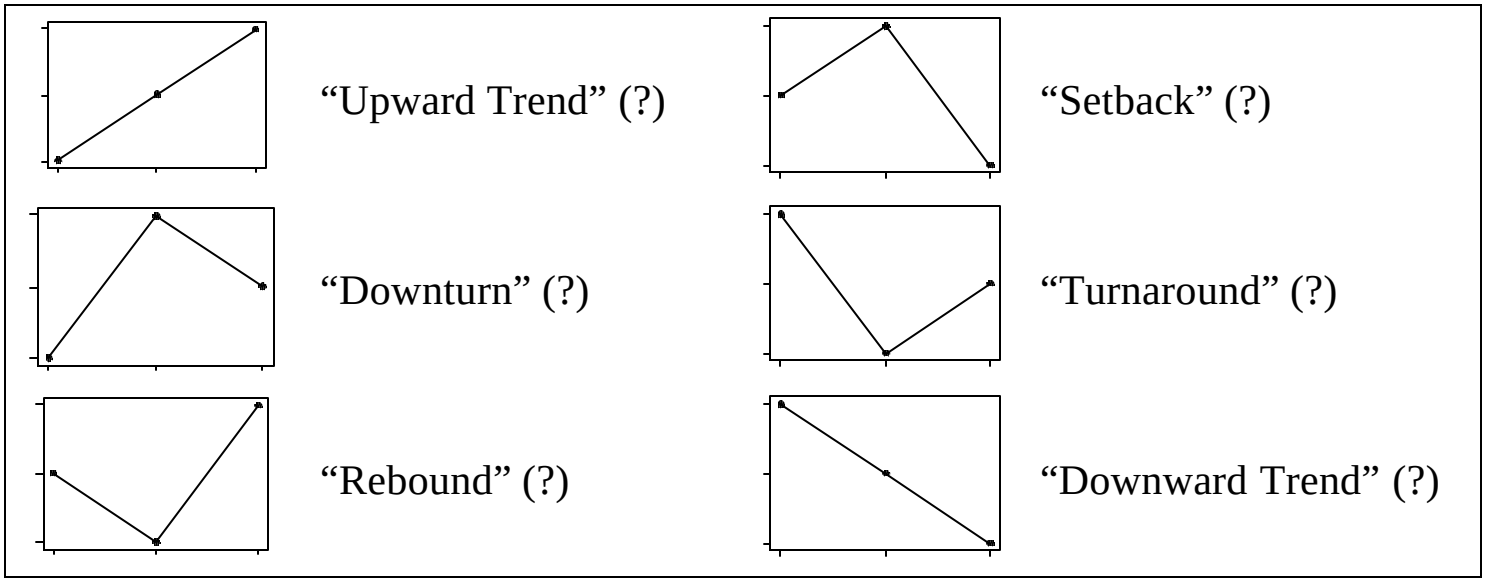
A further financial incentive:

a) If all three months meet the "stretch" goal, there will be a bonus,

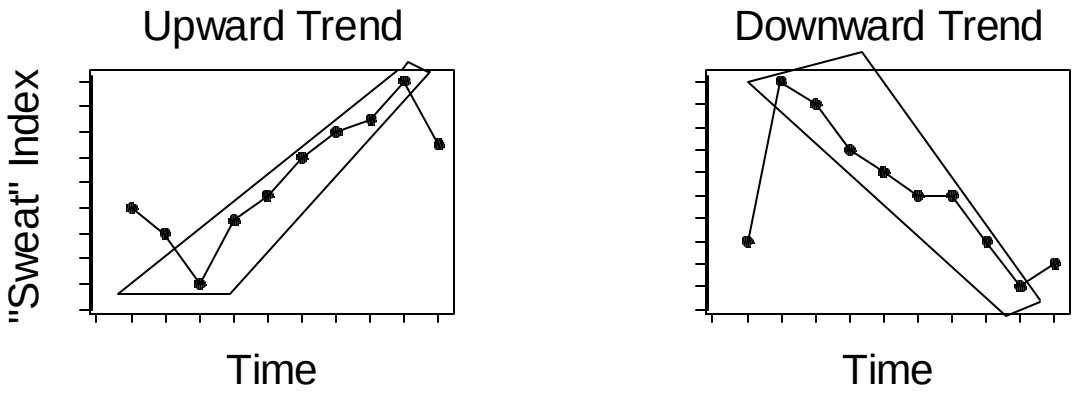
b) Any month in which the "Stretch" goal is bested by 25% will result in an incentive bonus provided the quarterly goal is met at a level of at least three times the monthly stretch goal as in (a).

❖ The "Myth of Trends"

Six Possible (& *RANDOM*) Sequences of Three Distinct Data Values



Statistical Representation of a Trend



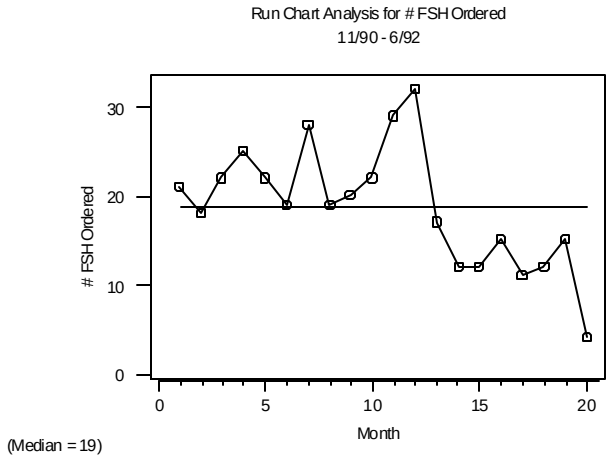
Special Cause – A sequence of SEVEN or more points continuously increasing or continuously decreasing – Indicates a trend in the process average.

Note 1: Omit entirely any points that repeat the preceding value. *Such points neither add to the length of the run nor do they break it.*

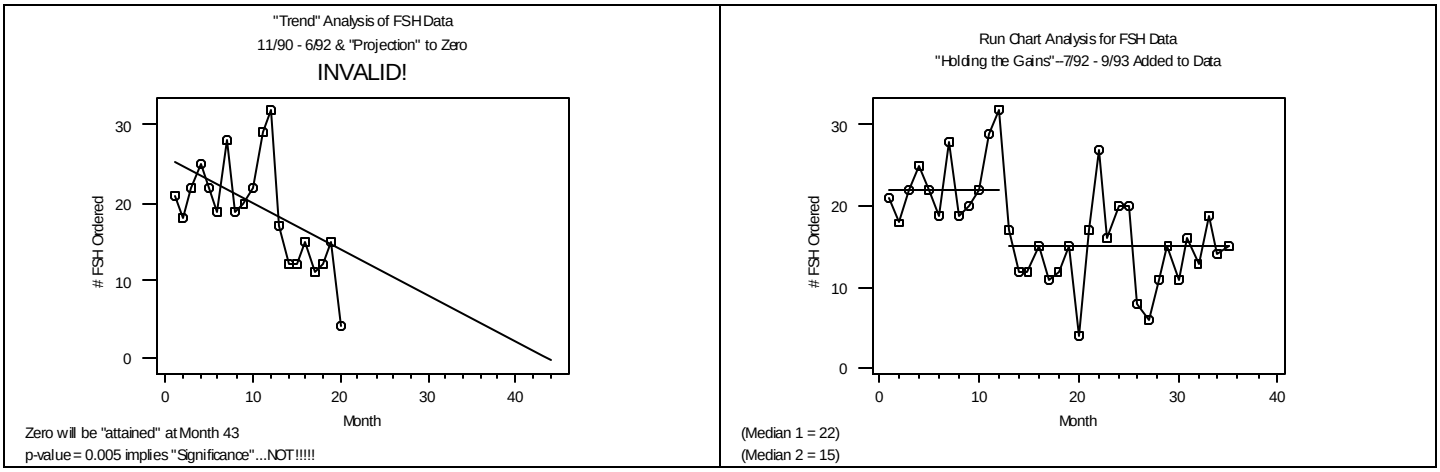
Note 2: If the total number of observations is ***20 or less***, **SIX** continuously increasing or decreasing points can be used to declare a trend.

This rule is to be used only when people are making conclusions from a tabulated set of data *without any context of variation for interpretation.*

- **More on trends: Fitting Inappropriate "Trend" Lines to a Time Series Data Sequence.**



Fitting an Inappropriate Trend Line to the Data
It is not uncommon for data like these (especially financial indices) to be analyzed via "trend" analysis with linear regression. From the p-value of 0.005, the regression is "obviously" significant.



❖ Old Habits Die Hard

Yet ANOTHER Meeting: "Statistical" Comparison of Three Hospitals' Lengths of Stay

The following data is handed out. You now have two hours…GO!

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean	Min	Max	Q1	Q3
LOS_1	30	3.027	2.900	3.046	0.978	0.178	1.000	4.800	2.300	3.825
LOS_2	30	3.073	3.100	3.069	0.668	0.122	1.900	4.300	2.575	3.500
LOS_3	30	3.127	3.250	3.169	0.817	0.149	1.100	4.500	2.575	3.750

Certainly, we have learned enough "tools" in school and during the CQI craze to do better than this!

What luck! You are "blessed"—A Six Sigma Black Belt is in your midst, and he brought his computer, onto which he loaded the raw data, to the meeting. Good news!--They ALL pass the normality test! Of course, we have to be cautious, as he appropriately warns us: *"Just because the data passes the test for normality doesn't necessarily mean that the data are normally distributed…only that, under the null hypothesis, the data cannot be proven to be non-normal."* **Got that?**

Hey…since the data can be assumed Normally distributed, one can proceed with the *analysis of variance* to generate the *95% confidence intervals*?

One-Way Analysis of Variance

```
Source     DF        SS        MS        F        P
Hospital    2     0.150     0.075     0.11    0.897
Error      87    60.036     0.690
Total      89    60.186
                                 Individual 95% CIs For Mean

Level       N      Mean     StDev  ----+---------+---------+---------+--
1          30    3.0267    0.9777  (--------------*--------------)
2          30    3.0733    0.6680     (--------------*--------------)
3          30    3.1267    0.8175       (--------------*--------------)
                                    ----+---------+---------+---------+--
Pooled StDev =   0.8307              2.80      3.00      3.20      3.40
```

"p-Value > 0.05: Therefore, no statistically significant difference as further confirmed by the overlapping 95% confidence intervals"

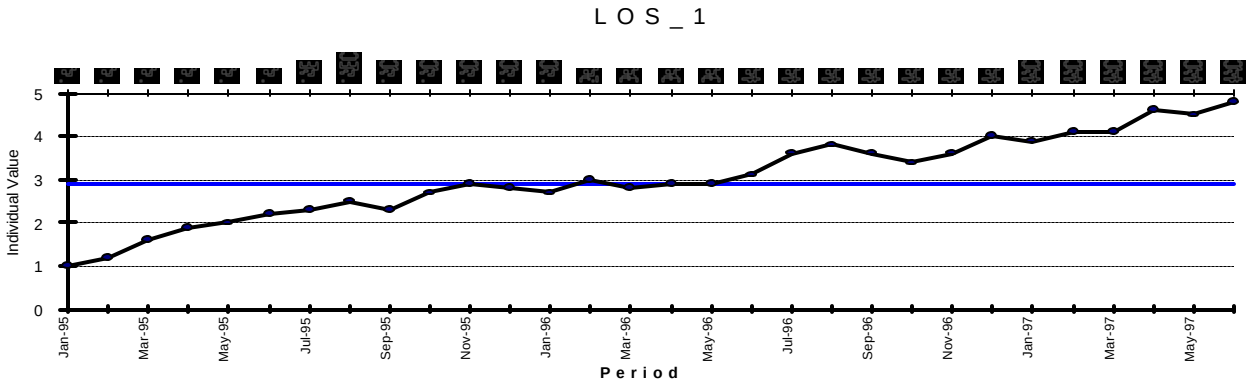
- **"He certainly sounds like he knows what he's talking about!"**

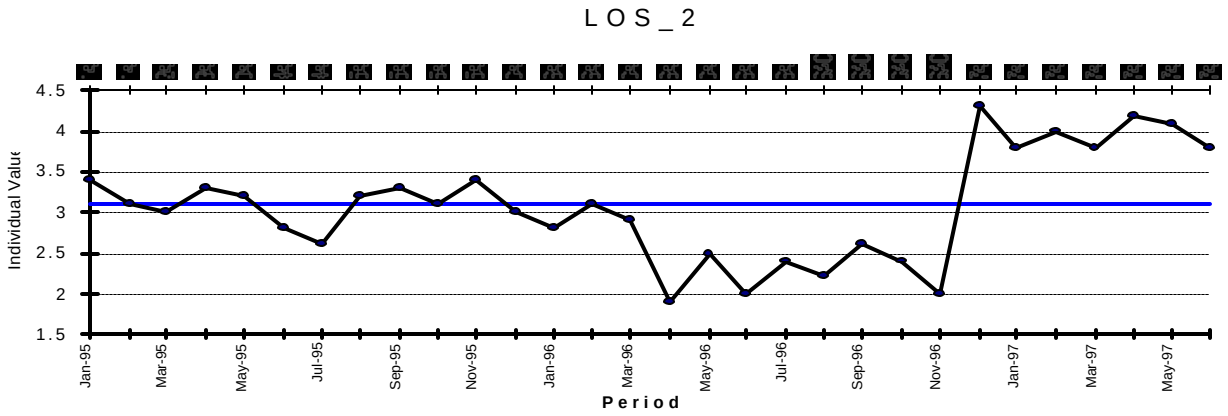
No jargon was left unturned:

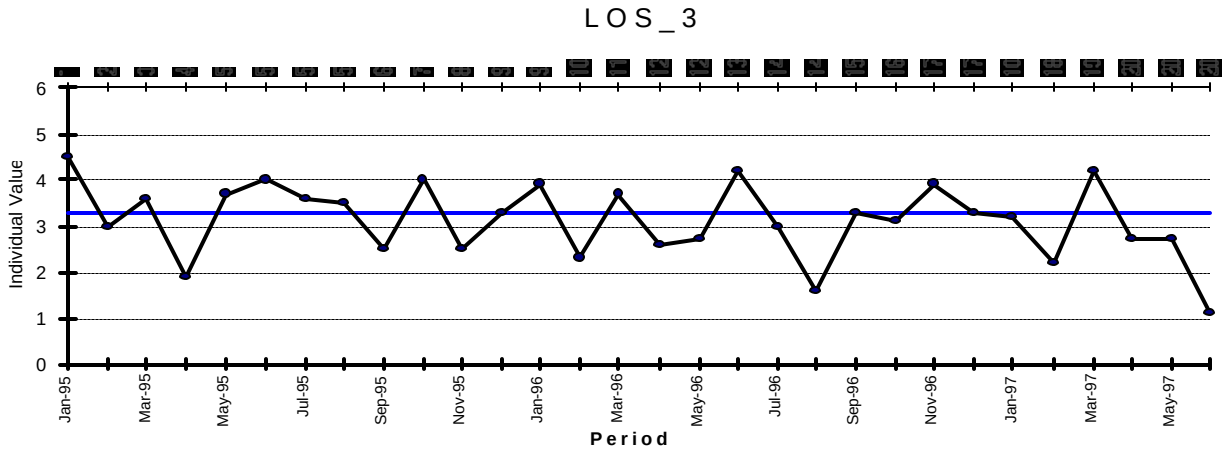
"Mean…Median…Standard Deviation…Normality…Histogram…p-value…Analysis of Variance (ANOVA)…95% Confidence Interval…Null Hypothesis…Statistical Significance…Standard Error of the Mean…F-test…Degrees of Freedom…"

How was this data collected?

*Was all this analysis **appropriate**?*







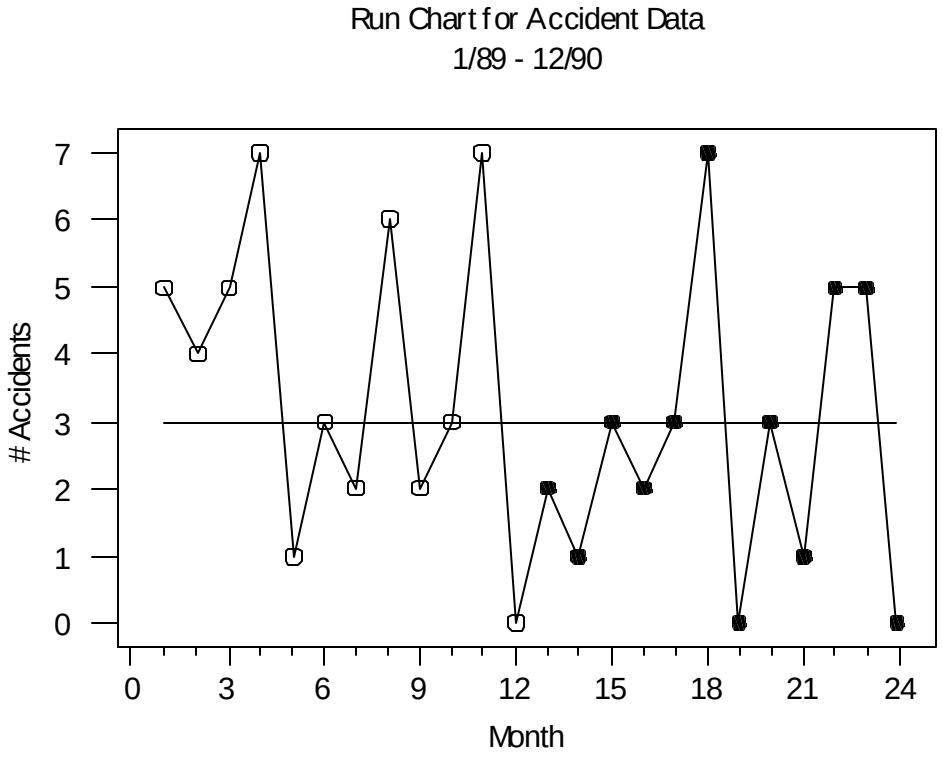
No Difference?!

[*"If I stick my right foot in a bucket of boiling water and my left foot in a bucket of ice water, on the average, I'm pretty comfortable."*]

"Plot the Bloody Dots!!!!!"

❖ Back to the Celebration Luncheon

The Safety Celebration Luncheon Data Run Chart



(Median = 3)

⇒ ***Is the process that produced the 32 the same as the process that produced the 45?***

If the conclusion that a significant 25% reduction in accidents occurred is correct, we would see either:

- A trend of seven going down from year 1 to year 2 and/or
- A run of length >= 8 above the median in year 1 and/or
- A run of length >= 8 below the median in year 2

Can you now summarize the results of the hard-working committee that concentrated on every single accident every month?

Matrix of Adverse Events

Event Type	Unit A	B	C	D	E	F	Total
1	0	0	1	0	2	1	4
2	1	0	0	0	1	0	2
3	0	**16**	1	0	2	0	**19**
4	0	0	0	0	1	0	1
5	2	1	3	1	4	2	**13**
6	0	0	0	0	3	0	3
27							
28					(less than 6 each)		
29							
Totals	6	**19**	7	3	**35**	7	77

❖ Back to the QA Report Meeting

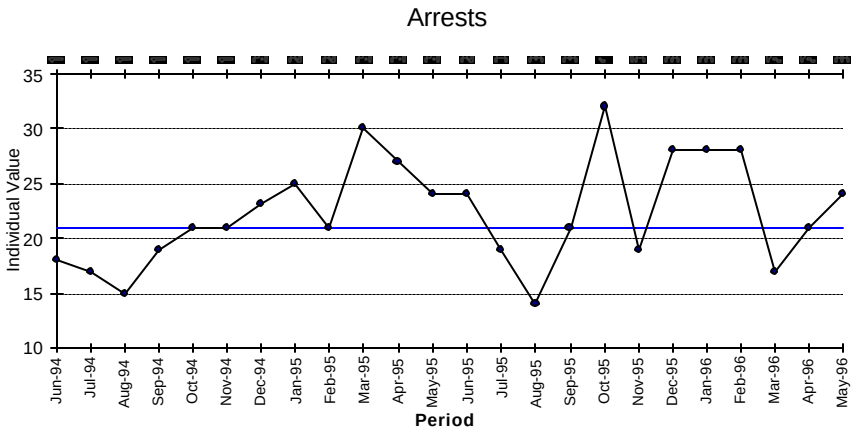
"We are running a slightly higher number of cardiac arrests per month. The total amount of cardiac arrests has risen from a mean of 21.75 (June 94- May 95), to 22.92 (June 95- May 96). This is an increase in 14 cardiac arrests in the last 12 months."

I think this is a fancy way of saying that 275 minus 261 equals 14! Is the conclusion above a reasonable one? The data for arrests, shown below, was sorted to calculate the median, then used to "plot the dots."

Arrests	Mo/Year	Sorted
18	6/94	14
17	7/94	15
15	8/94	17
19	9/94	17
21	10/94	18
21	11/94	19
23	12/94	19
25	1/95	19
21	2/95	21
30	3/95	21
27	4/95	21
24	5/95	**21**
24	6/95	**21**
19	7/95	23
14	8/95	24
21	9/95	24
32	10/95	24
19	11/95	25
28	12/95	27
28	1/96	28
28	2/96	28
17	3/96	28
21	4/96	30
24	5/96	32

Median = (21 + 21) /2

= 21



⇒ **Is the *process* that produced the 275 is the same as the *process* that produced the 261?**

Since the analyst's theory is that there has been an increase:

- **Do you see a trend of 7 going up?**

Count the run lengths (Ignore points literally on the median):

- **Do you have a run length >= 8 *below* the median in the 1st year and/or…**
- **a run length >= 8 *above* the median in the 2nd year?**

"Next we interpreted the data relating to Vfib Cardiac Arrests…This could be significant to our outcome, and…indicates a need for more sophisticated statistical analysis. It was already shown that the number of cardiac arrests has increased by a mean of 1.17 per month. Now we are adding to that increase, a decrease of times we are seeing Vfib as the initial rhythm. From June 1994 to May 1995 we arrived on scene to find Vfib as the initial rhythm with an overall mean of 6.75 times. That gave us a capture rate of 32.03%. This last year, June 1995 - May 1996, we are arriving to find Vfib as the initial rhythm with an overall mean of 5.92, and a capture rate of 25.81%. This obviously means that over the last year, we have responded to more cardiac arrests and found them in more advanced stages of arrest."

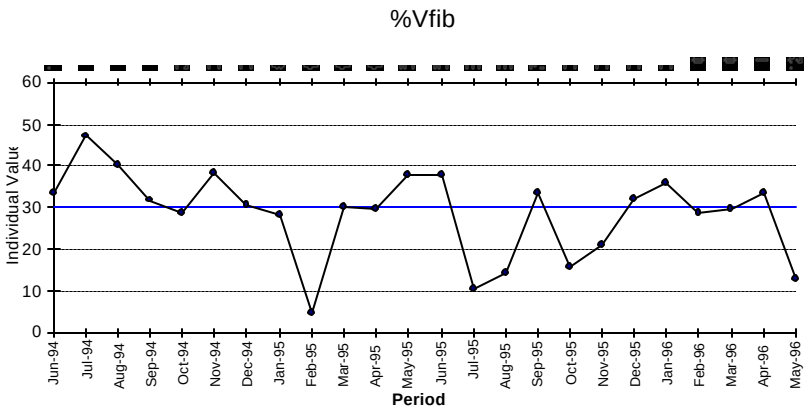
"Don't just do something…STAND there"…and "plot the dots." But WHICH dots?

It seems that there is interest in the *proportion of arrests that are in Vfib.* So, you pull out your calculator, calculate the %Vfib arrests for each month (For example, 6/94 is 6 out of 18 or 33.33%), and do a run chart of the percentages [Shown below].

Arrests	Vfib	%Vfib	Mo/Year	%Vfib_Sort
18	6	33.33	6/94	4.76
17	8	47.06	7/94	10.53
15	6	40.00	8/94	12.50
19	6	31.58	9/94	14.29
21	6	28.57	10/94	15.63
21	8	38.10	11/94	21.05
23	7	30.43	12/94	28.00
25	7	28.00	1/95	28.57
21	1	4.76	2/95	28.57
30	9	30.00	3/95	29.41
27	8	29.63	4/95	29.63
24	9	37.50	5/95	**30.00**
24	9	37.50	6/95	**30.43**
19	2	10.53	7/95	31.58
14	2	14.29	8/95	32.14
21	7	33.33	9/95	33.33
32	5	15.63	10/95	33.33
19	4	21.05	11/95	33.33
28	9	32.14	12/95	35.71
28	10	35.71	1/96	37.50
28	8	28.57	2/96	37.50
17	5	29.41	3/96	38.10
21	7	33.33	4/96	40.00
24	3	12.50	5/96	47.06

Median = (30 + 30.43)/2
= 30.215

Tot_Arr	Tot_Vfib	%Vfib	Period
261	81	31.03	6/94-5/95
275	71	25.82	6/95-5/96

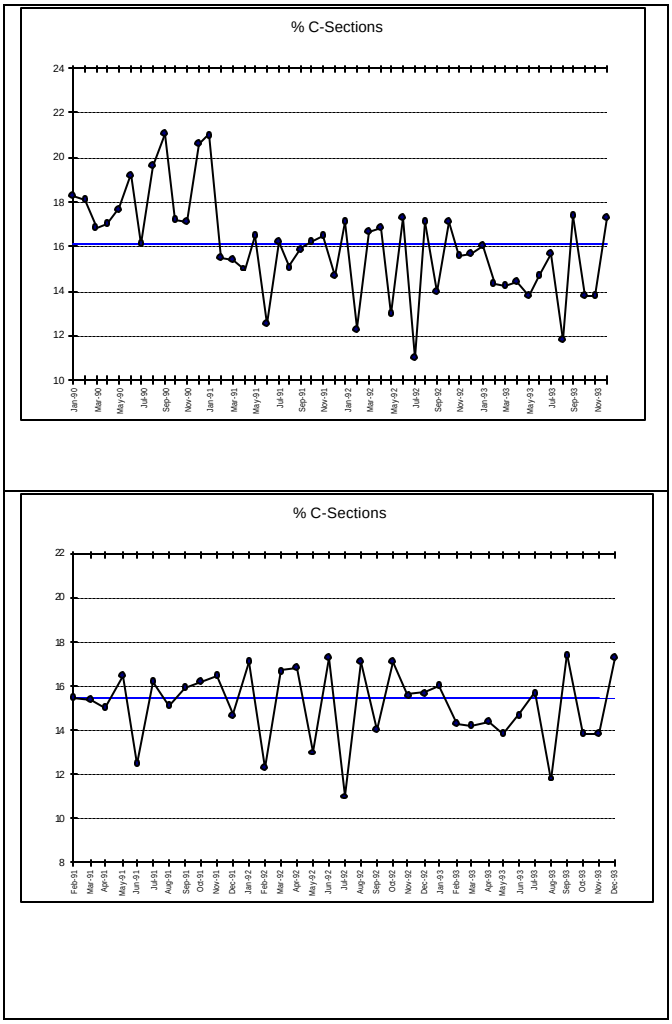


⇒ **Is the process that produced the 25.82% is the same as the process that produced the 31.03%?**

- **You know the drill…and you conclude?**

❖ Back to the C-Section Meeting:

Month	C-Sec	Births	%C-Sec
Jan 95	47	257	18.29
Feb	48	265	18.11
Mar	50	298	16.78
Apr	50	294	17.01
May	62	351	17.66
Jun	47	245	19.18
Jul	54	336	16.07
Aug	55	281	19.57
Sept	62	294	21.09
Oct	49	285	17.19
Nov	46	269	17.10
Dec	51	247	20.65
Jan 96	60	285	21.05
Feb	40	258	15.50
Mar	43	280	15.36
Apr	44	293	15.02
May	57	346	16.47
Jun	38	304	12.50
Jul	56	346	16.18
Aug	47	311	15.11
Sept	44	276	15.94
Oct	46	283	16.25
Nov	44	267	16.48
Dec	39	265	14.72
Jan 97	50	292	17.12
Feb	46	373	12.33
Mar	45	269	16.73
Apr	48	285	16.84
May	39	300	13.00
Jun	50	289	17.30
Jul	31	282	10.99
Aug	47	275	17.09
Sept	42	299	14.05
Oct	49	286	17.13
Nov	38	244	15.57
Dec	39	249	15.66
Jan 98	38	237	16.03
Feb	32	224	14.29
Mar	37	260	14.23
Apr	40	277	14.44
May	41	296	13.85
Jun	40	272	14.71
Jul	42	267	15.73
Aug	34	288	11.81
Sept	42	241	17.43
Oct	37	268	13.81
Nov	32	232	13.79
Dec	40	231	17.32



Is there "hidden opportunity?"—Revisiting the individual physician data

Here is the *actual* data for each of the 15 physicians who practice at the hospital.

MD	C-Sections	Deliveries	%C-Sections
1	27	317	8.52
2	29	337	8.61
3	36	356	10.11
4	36	309	11.65
5	36	266	13.53
6	34	233	14.59
7	30	196	15.31
8	57	353	16.15
9	35	215	16.28
10	47	288	16.32
11	59	356	16.57
12	63	350	18.00
13	68	350	19.43
14	84	401	20.95
15	52	241	21.58
Total	**693**	**4568**	**15.17%**

(Std. Dev. = 4.09)

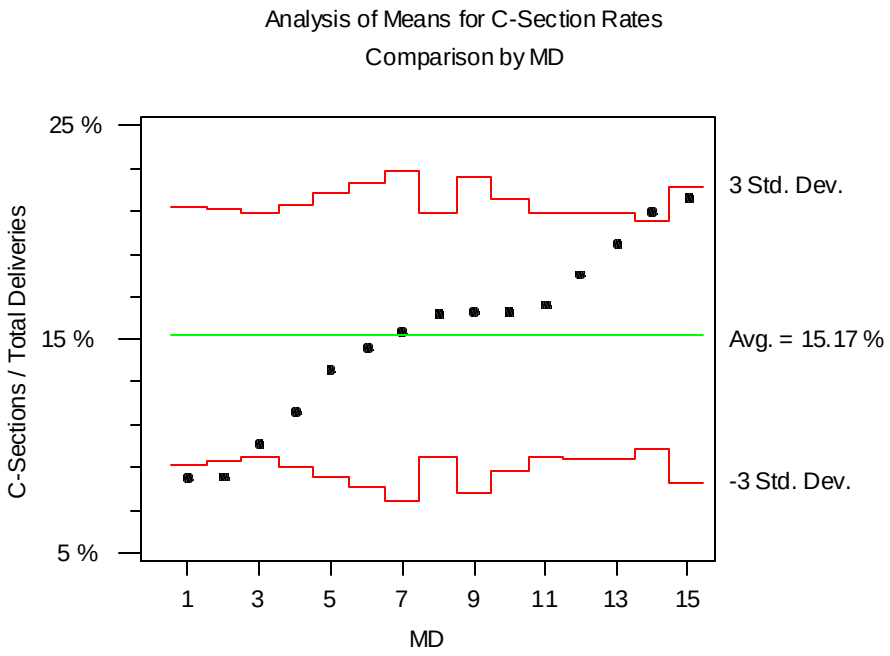
Only because the 2nd run chart exhibits "stable" behavior, *it is now appropriate to aggregate the data,* **for this stable time period only**, *and stratify it by physician (common cause strategy).*

There is a tendency to use some naïve (and inappropriate) testing of the data for "normality" along with a threshold of "one or two standard deviation" limits based on the traditional calculation of sigma, $\sqrt{\frac{1}{n-1}\Sigma\left(X_i - \overline{X}\right)^2}$. Applying this technique to the individual 15 percentages yielded a "normal distribution" and a standard deviation of 4.09. ***Wrong!***

How about using an analysis *appropriate for the way the data were collected*? The "system" value is determined as the total number of C-Sections by the 15 physicians divided by the *total number* of deliveries performed by the 15 physicians. ($(693/4568) = 15.17\%$). We are going to assume that ALL physicians' C-Section rates are 15.17% *unless the data tell us otherwise.*

Based on the statistical theory appropriate for this percentage data, standard deviations must be calculated separately ***for each physician*** because each performed a different number of deliveries. The formula is

$$\sqrt{\frac{(15.17)x(100-15.17)}{\text{Total deliveries by MD}}}\ .$$

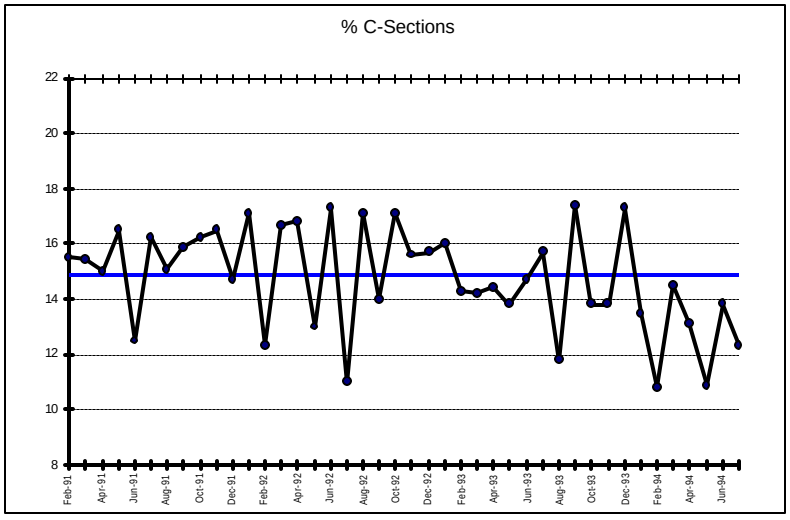


Common Cause Range: $15.17 \pm 3 \times \sqrt{15.17 \times (100 - 15.17) / (\text{Total Deliveries by MD})}$

⇒ **Suppose a goal had been set for "no more than 15% C-sections?" Who should get feedback?**

⇒ **How about a "stretch" goal of 12% for next year?**

Now suppose there had been a discussion among the physicians about their different processes for handling labors. The subsequent seven months had C-section rates of 13.53, 10.85, 14.5, 13.12, 10.94, 13.85, and 12.31 %. These have been added on to the previous run chart [new median of 14.785 calculated by adding these 7 numbers to the previous most stable system —"Innocent until proven guilty"] and are shown below:



Six Common Statistical Traps

TRAP	***PROBLEM***	***COMMENT***
Trap 1: Treating **all** observed variation in a time series data sequence as special cause.	Most common form of "Tampering"—treating common cause as special cause.	Given two numbers, one will be bigger! Very commonly seen in traditional monthly reports: Month-to-Month comparisons; Year-Over-Year plotting and comparisons; Variance reporting; Comparisons to arbitrary numerical goals.
Trap 2: Fitting inappropriate "trend" lines to a time series data sequence.	a) Another form of "Tampering"—attributing a specific type of special cause (linear trend) to a set of data which contains only common cause. b) Attributing an *inappropriate* specific special cause (linear trend) to a data time series that contains a different kind of special cause.	Typically occurs when people always use the "trend line" option in spreadsheet software to fit a line to data with no statistical trends. Improvement often takes place in "steps," where a stable process moves to a new level and remains stable there. However, a regression line will show statistical significance, implying that the process will continually improve over time.
Trap 3: Unnecessary obsession with and incorrect application of the Normal distribution.	a) A case of "reverse" tampering—treating special cause as common cause. b) Inappropriate routine testing of all data sets for Normality.	Ignoring the time element in a data set and inappropriately applying techniques based on the Normal distribution can cause misleading estimates and inappropriate predictions of process outputs. Misapplying Normal distribution theory and "traditional" calculations to percentage or count data.
Trap 4: Incorrect calculation of standard deviation and "sigma" limits.	Since much improvement comes about by exposing and addressing special cause opportunities, the traditional calculation of standard deviation can typically yields a grossly inflated variation estimate.	Because of this inflation, people have a tendency to arbitrarily change decision limits to two (or even one!) standard deviations from the average or "standard". Using a *three* standard deviation criterion with the *correctly* calculated value of sigma gives approximately an *overall* statistical error risk of 0.05.
Trap 5: Choosing **arbitrary** cutoffs for "above" average and "below" average. [Ranks, Quartiles]	There is actually a "dead band" of common cause variation on either side of an average that is determined from the data themselves.	Approximately half of a set of numbers will naturally be either above or below average. Potential for tampering appears again. Percentages are especially deceptive in this regard.

Trap 6: Improving processes through the use of arbitrary numerical goals and standards.	Any process output has a natural, inherent capability within a common cause range. It can perform only at the level its inputs will allow.	Goals are merely wishes regardless of whether they are necessary for survival or arbitrary. Data must be collected to assess a process's natural performance relative to a goal.

Lessons learned

1. Make sure that any data being used were collected *specifically* for the current purpose,
2. Understand *how* the numbers were calculated and how the data were *collected*,
3. Make sure any analysis is *appropriate* for the way the data were collected,
   (a) Tables of raw numbers, summary "stats," and bar graph presentations are virtually worthless,
   (b) The "normal" distribution is highly overrated and very rarely used in improvement,
   (c) "Traditional" calculation of the standard deviation will typically yield an inflated estimate,
4. "Plotting the dots" of an indicator over time is a powerful but simple method for studying a process,
5. Arbitrary numerical goals and percentages to declare "outliers" are inappropriate,
6. Reacting to individual data points in a "stable" system is a *no yield* strategy,
7. A "stable" system can be "dissected" statistically to look for hidden opportunities,
8. Percentages must be interpreted in the context of the denominators producing the percentages,
9. Looking at the pattern of "dots" after an intervention can be used to assess the affect of the intervention.

Useful References

Balestracci D. "Data 'Sanity': Statistical Thinking Applied to Everyday Data," Special Publication of the American Society for Quality Statistics Division, Summer 1998. Available through the Quality Information Center by calling (800)-248-1946 ($15). [An executive summary of the "traps"]

Balestracci D & Barlow J. *Quality Improvement: Practical Applications for Medical Group Practice, 2nd Edition*. Englewood, CO. Center for Research in Ambulatory Health Care Administration (CRAHCA), 1996. Phone: (303)-397-7888 to order ($59).

Executive Learning Inc. Brentwood, TN: Executive Learning Inc. [800-929-7890]

Hacquebord, H. "Health care from the perspective of a patient: Theories for improvement." *Quality Management in Health Care*. 1994, 2(2). [An personal analysis of the "back surgery from hell"]

Joiner B. *Fourth Generation Management: The New Business Consciousness*. New York, NY: McGraw-Hill, Inc., 1994.

Wheeler DJ.

Understanding Variation: The Key to Managing Chaos. Knoxville, TN: SPC Press, Inc., 1986.

Building Continual Improvement: A Guide for Business. Knoxville, TN: SPC Press, Inc., 1998.

Classics from Donald J. Wheeler. Knoxville, TN: SPC Press, Inc., 1999.