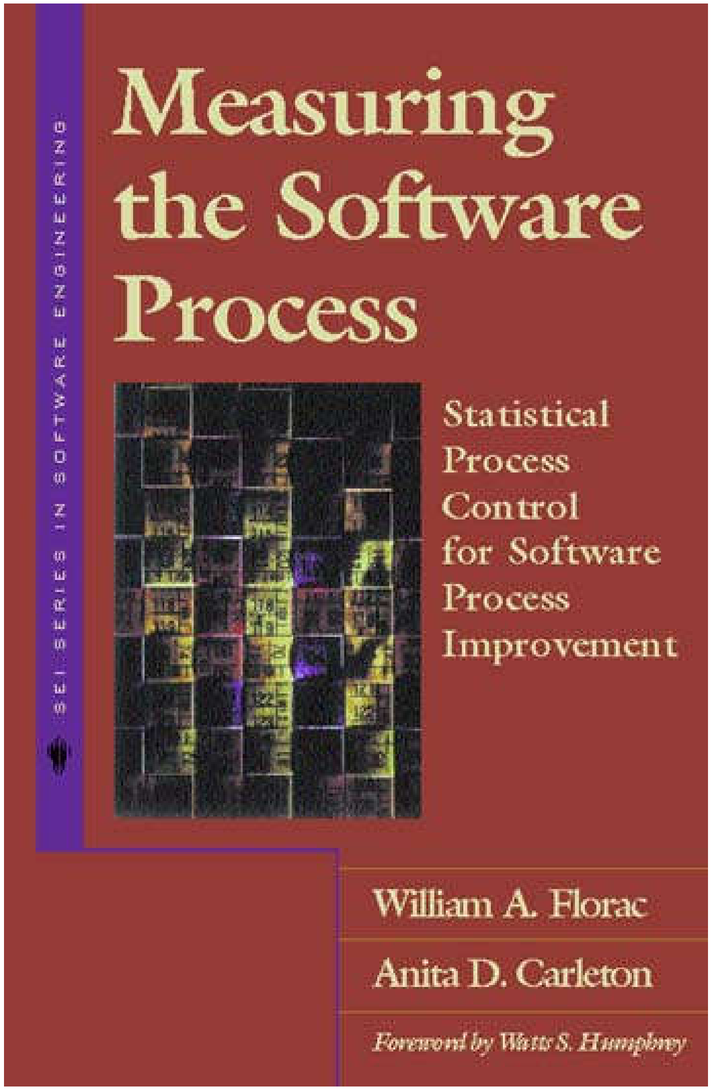
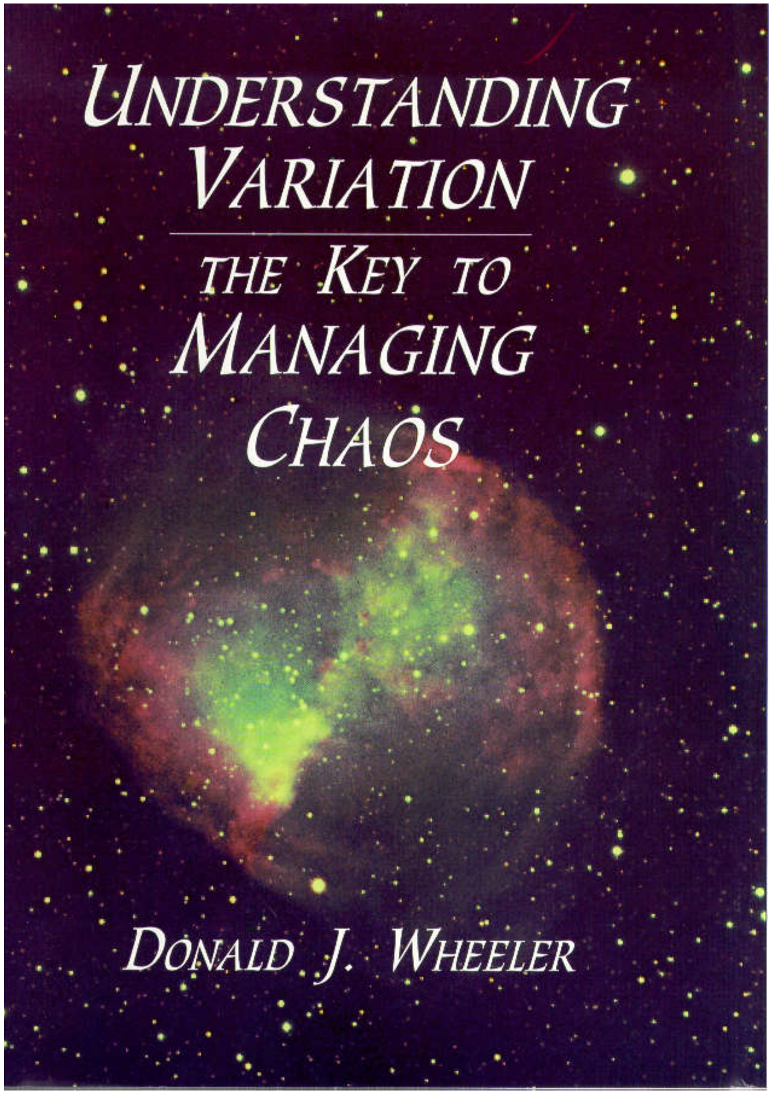
Statistical Thinking

Dr. Mark C. Paulk

SE 6388, Software Project Planning and Management

Recommended Textbooks





Process Thinking

"<u>Process approach</u>: A desired result is achieved more efficiently when activities and related resources are managed as a process."

- **ISO 9004:2000**

Key aspects of process thinking:

- **all work is a series of interconnected processes**
- **the majority of the problems are in the process**
- **managers must focus on fixing processes, not blaming people**
- **managers must ensure that the needed processes are in place**
  - **Hare, Hoerl, Hromi, and Snee, "The Role of Statistical Thinking in Management," ASQC Quality Progress, February 1995.**

Systems Thinking

"<u>System approach to management</u>: Identifying, understanding, and managing interrelated processes as a system contributes to the organization's effectiveness and efficiency in achieving its objectives."

- **ISO 9004:2000**

Two basic options for closing a performance gap:

- **Work Harder (a balancing feedback loop)**
- **Work Smarter (improve the capability of the process)**

Working Smarter has its limitations… there is often a substantial delay between investing in improvement and reaping the benefits.

- **Repenning and Sterman, "Nobody Ever Gets Credit for Fixing Problems that Never Happened," California Management Review, Summer 2001.**

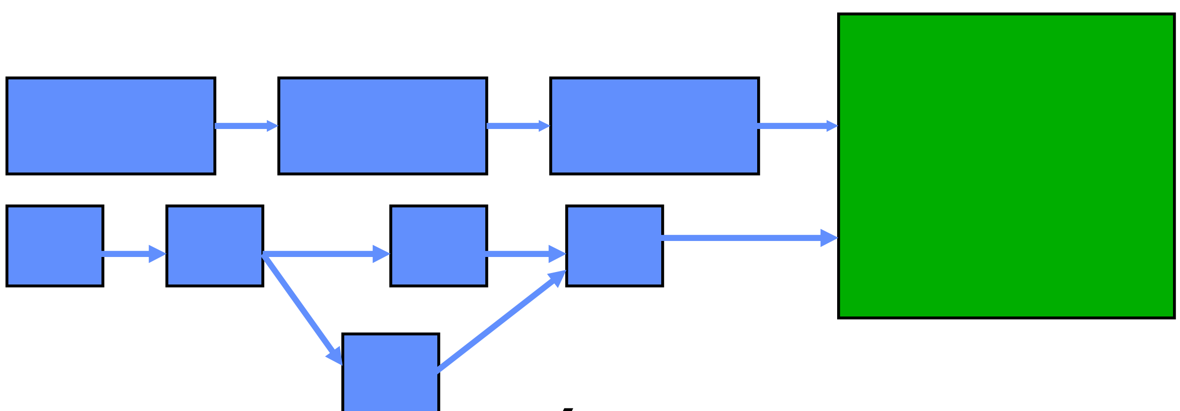
Lean thinking – minimize waste

Systems Level Perspective

Avoid suboptimization by balancing process management with a systems perspective.

Optimal processes at the (sub)process level will not lead to an optimal solution at the project level!

- **non-intuitive, but well-known, observation from industrial engineering**



Statistical Thinking

All work is a series of interconnected processes.

All processes are variable.

Understanding variation is the basis for management by fact and systematic improvement.

CMMI, OID Tip: Changes must be measurably better…

Repetitive vs Repeatable Processes

The software process is performed by people, not machines.

The software process is (or can be) repeatable, but not repetitive.

The act of measuring and analyzing will change behavior – potentially in dysfunctional ways.

Profound Knowledge of the Domain

Prerequisites for statistical analysis

- **“homogenous” subgroups**
- **“random” sampling**

Comparing apples to apples, oranges to oranges

Software CMM v2C, Level 4 KPA on product lines, reuse, etc.

CMMI, OPP SP1.4(i) – subgroups of the organization

- **product line, line of business, application domain, …**

The analyst must understand the domain to make informed decisions…

Evidence-Based Management

“<u>Factual approach to decision making</u>: Effective decisions are based on the analysis of data and information.”

- **ISO 9004:2000**

“It isn’t what we don’t know that gives us trouble, it’s what we know that ain’t so.”

- **Will Rogers**

Quality Culture

Culture – how we do things around here

Institutionalization – how we inform, support, and reinforce the way we want to do things

Professionalism – accepted practice about the best way of doing things as a discipline

Software CMM, PC.GO.2 organization-wide participation

CMMI, OID Tip: everyone in the organization should be empowered and encouraged to suggest potential improvements.

"<u>Involvement of people</u>: People at all levels are the essence of an organization and their full involvement enables their abilities to be used for the organization's benefit."

- **ISO 9004:2000**

Operational Definitions

Two criteria: ***communication*** **and** ***repeatability***

How is a line of code defined?
- ***new, modified, deleted, retained function***

What are the time units?
- ***hours vs minutes***

How is a defect defined?
- ***severity, criticality, impact***

Is the data collected at the same point in the process each time?
- ***before or after compile***

A Measurement Example

How many lines of code do we have here?

```
# include <stdio.h>
# include <stdlib.h>

// This program prints “Hello world” to the screen.

main()
{

    printf(“Hello world“);

}
```

"An operational definition [is one] which reasonable men can agree on and do business with."

"Shewhart believed his work on operational definitions to have been of greater importance than his development of the theory of variation and of the control chart."

"There is no true value of anything."

Chapter 7 in Henry R. Neave, The Deming Dimension.

SEI Core Measures

Checklist-based approach with strong emphasis on operational definitions

Measurement areas where checklists have already been developed include:

- **effort**
- **size**
- **schedule**
- **quality**

See http://www.sei.cmu.edu/sema/publications.html for relevant SEI technical reports.

Example of an Aid for Operational Definitions

Problem Status	**Include**	**Exclude**	**Value Count**	**Array Count**
Open	🗐		🗐	
Recognized				🗐
Evaluated				🗐
Resolved				🗐
Closed	🗐		🗐	
Problem Type	**Include**	**Exclude**	**Value Count**	**Array Count**
Software defect				
Requirements defect	🗐		🗐	
Design defect	🗐		🗐	
Code defect	🗐		🗐	
Operational document defect	🗐		🗐	
Test case defect		🗐		
Other work product defect		🗐		
Other problems				
Hardware problem		🗐		
Operating system problem		🗐		
User mistake		🗐		
Operations mistake		🗐		
New requirement/enhancement		🗐		
Undetermined				
Not repeatable/Cause unknown		🗐		
Value not identified		🗐		
Uniqueness	**Include**	**Exclude**	**Value Count**	**Array Count**
Original	🗐			
Duplicate		🗐	🗐	
Value not identified		🗐		
Criticality	**Include**	**Exclude**	**Value Count**	**Array Count**
1st level (most critical)	🗐			🗐
2nd level	🗐			🗐
3rd level	🗐			🗐
4th level	🗐			🗐
5th level	🗐			🗐
Value not identified		🗐		
Urgency	**Include**	**Exclude**	**Value Count**	**Array Count**
1st (most urgent)	🗐			
2nd	🗐			
3rd	🗐			
4th	🗐			
Value not identified		🗐		

Process Consistency

Operationally different processes – discovered after data analysis begins…

Are "pre-reviews" performed (on some modules)?

Are inspection preparation and meeting rates consistent?

- ***across various sizes of work product?***

Are inspection preparation and meeting rates reasonable?

- ***follow rules for performing reviews correctly?***

Are some reviewers required? Optional?

Aggregation

Sources of variation can be overlooked when data is highly aggregated.

- **poor operational definitions**
- **working with data whose elements are combinations (mixtures) of values from different sources**
- **inadequate contextual information**
- **lack of traceability from data back to its original context**

Statistical Process Control

The use of *statistical* tools and techniques

to analyze a *process*
or its *outputs*

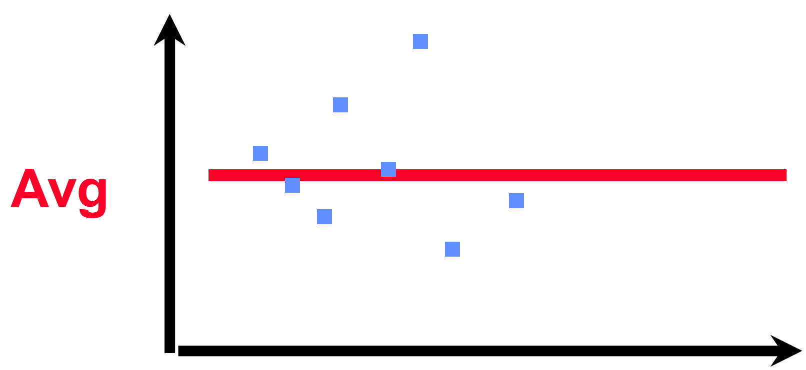
to control, manage, and improve
the *quality* of the output
or the *capability* of the process.

Statistical process control can, in a sense, be considered the effective use of statistical tools (some of which may not even be quantitative).

Why Use Statistics?

Most software organizations use the average of previous performance in estimating, planning, and making commitments.

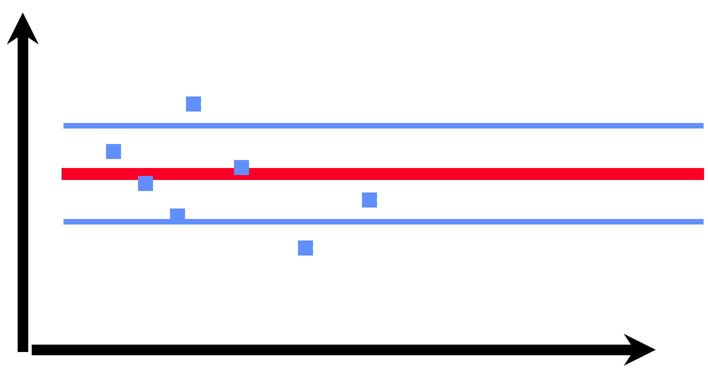
- **average is a statistic – a measure of central tendency (mean, mode, median)**



More Than "Average"

Software organizations usually establish performance targets for what to expect and where to do better.

Intervals may be preferred over "point" targets.



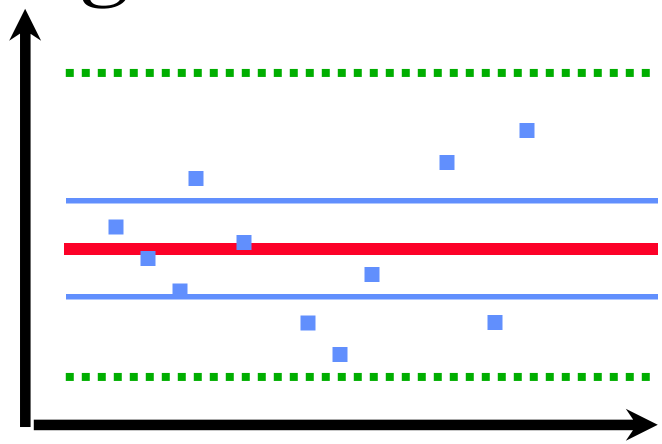
Avg +/- 20%
("Voice of Customer")

"Voice of the Customer" describes the performance we want to achieve.

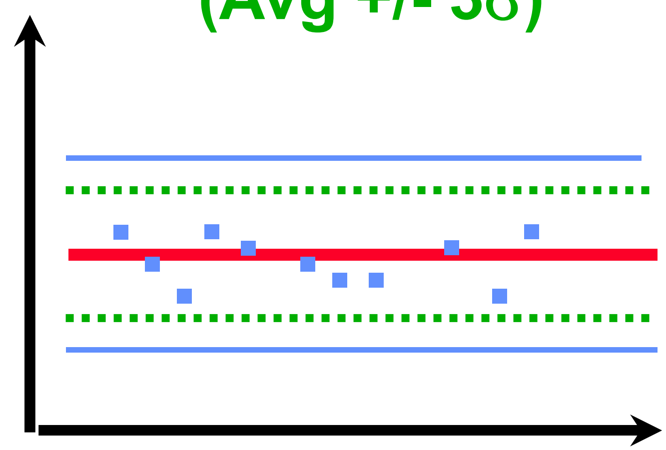
Statistics Can Lead To Understanding

Knowing what is <u>possible</u> with the current process may indicate the kind of action necessary to achieve those targets!

"Voice of the Process" describes the expected performance when making *realistic* commitments.



"Voice of the Process" (Avg +/- 3σ)



Central Tendency and Dispersion

Central tendency **implies location**

- **“middle” of a group of values**
- **“balance” point**
- ***examples: mean, median***

Dispersion **implies spread**

- **distance between values**
- **how much the values tend to differ from one another**
- ***examples: range, (sample) standard deviation***

Used together to understand the likelihood of a particular value

"Parameters" and "Statistics"

***Parameter*: A measurable characteristic about an entire population.**

- ***examples: mean (μ), population standard deviation (σ), variance (σ^2)***

***Statistic*: A measurable characteristic about a sample of a population.**

- ***examples: average ($\bar{x}$, X-bar, X_{bar}), sample standard deviation (s)***

Parameters of statistical distributions are frequently estimated by statistics

- $\hat{\mu} = \bar{x}$
- $\hat{\sigma} = s$

Distributions

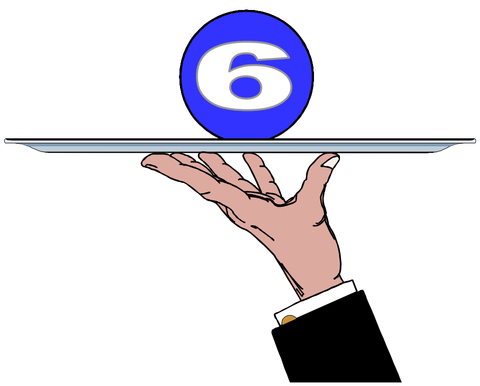
Populations of data are characterized as distributions in most statistical procedures.

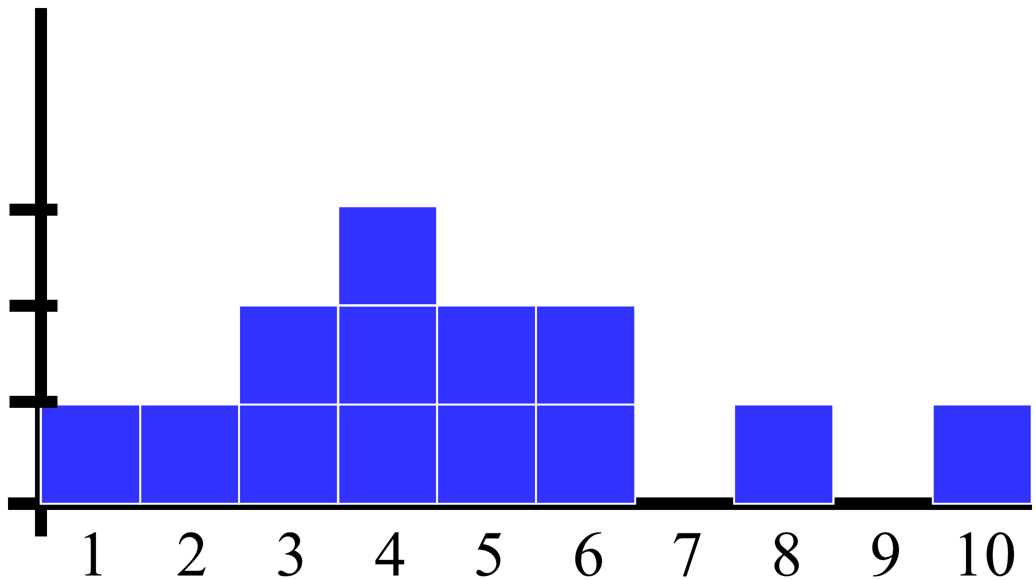
- **expressed as an “assumption” for the procedure**
- **can be represented using an equation**

Examples of distributions you may come across in SPC include

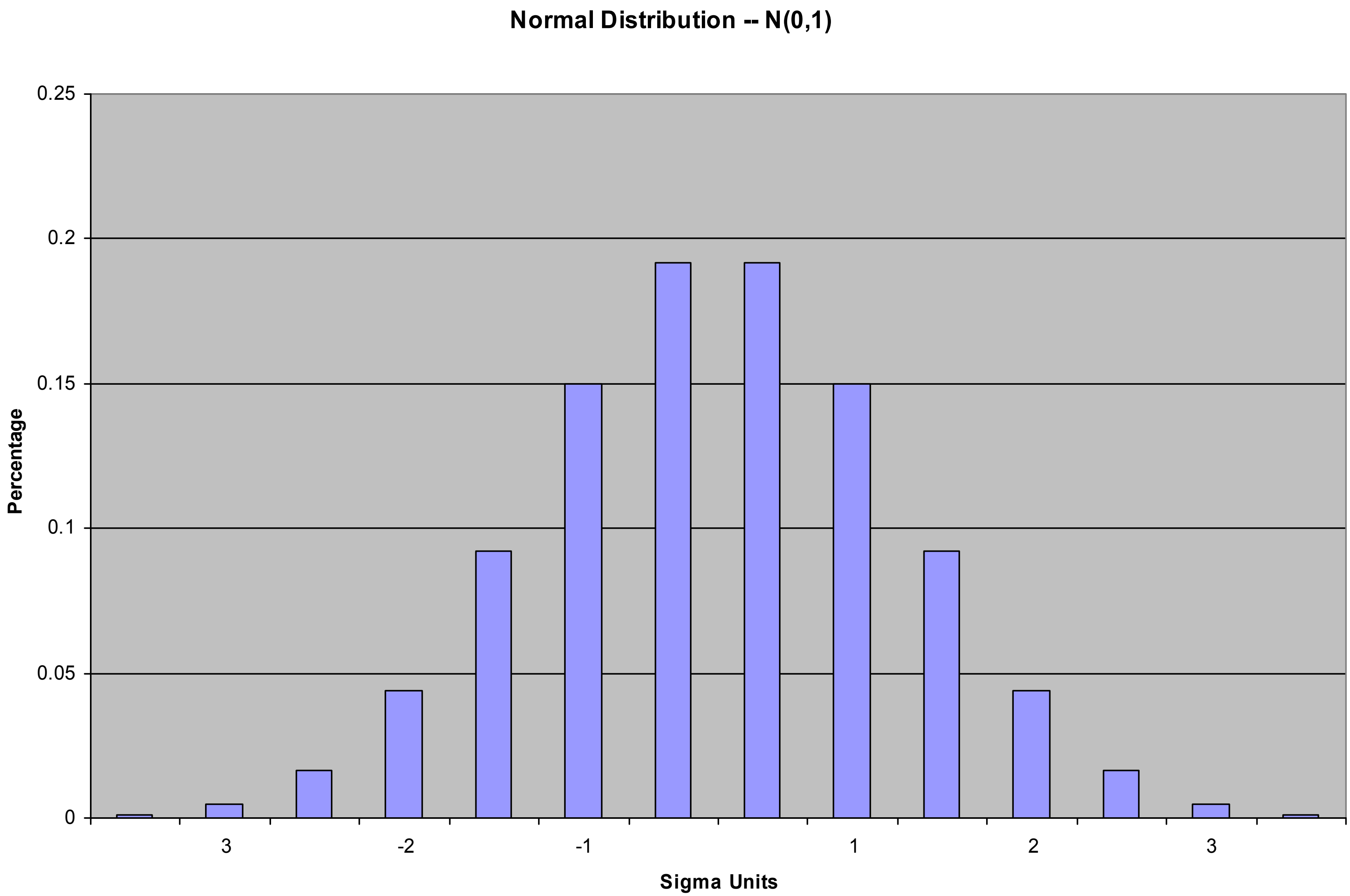
- **(standard) normal (aka Gaussian)**
- **binomial**
- **Poisson**

How Distributions Are Formed



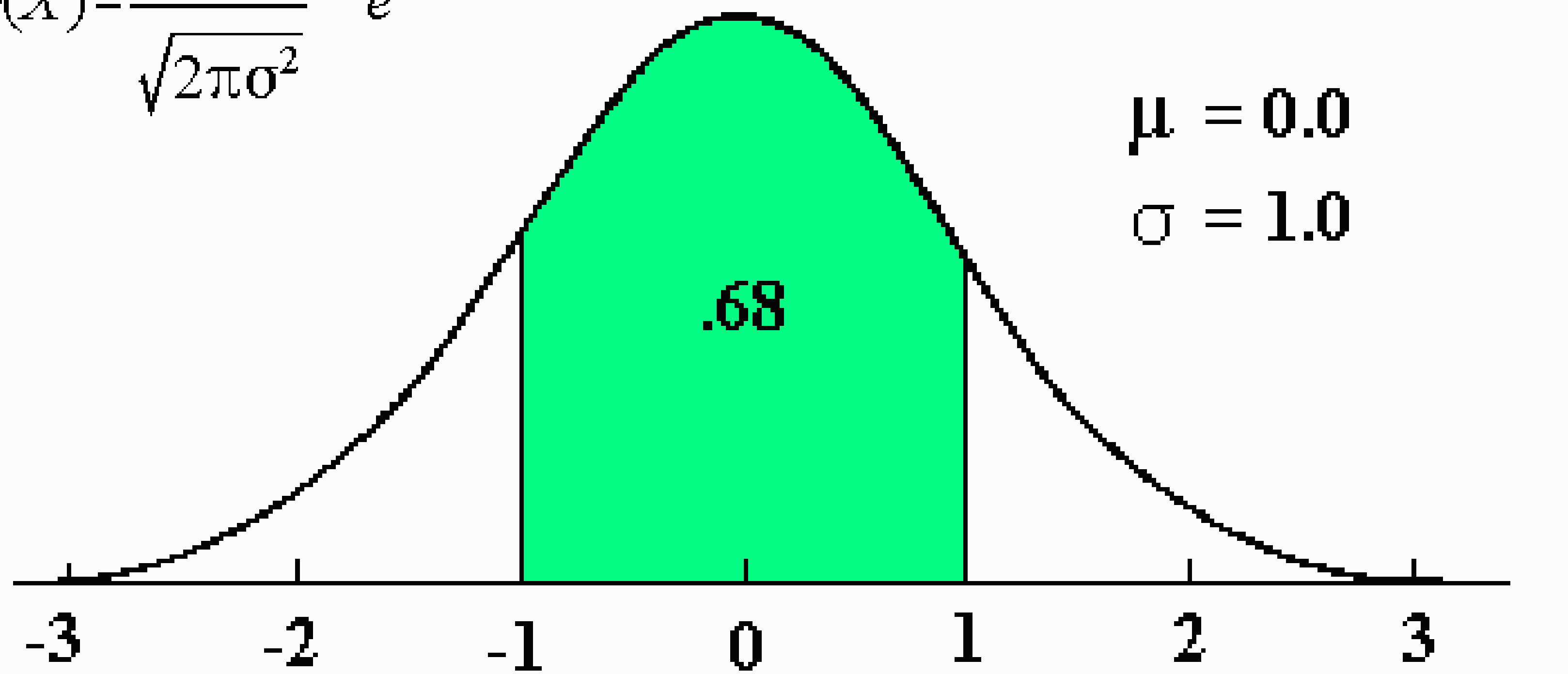


What Is A Normal Distribution?



The Normal Distribution

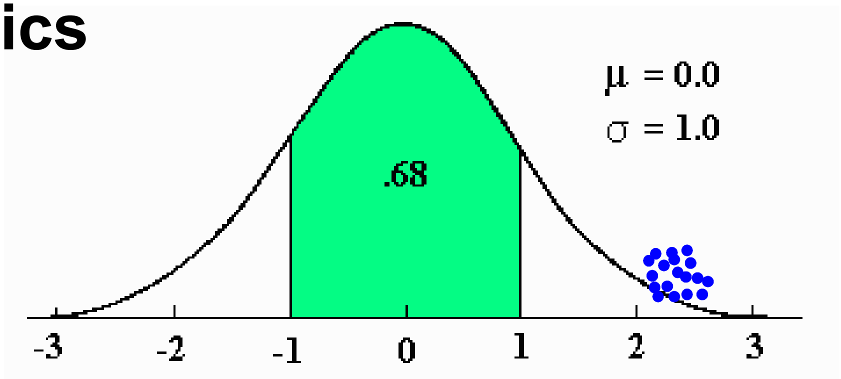
$$p(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{\frac{-(X-\mu)^2}{2\sigma^2}}$$



Use of Distributions in Statistics

Statistical techniques frequently use distributions to characterize the probability of observations.

- **for a normal distribution, 68% of the values will fall within 1 standard deviation**
- **a sample comprised only of values beyond 2 standard deviations is not very likely in this case**
- **this likelihood can be quantified using statistics**



Binomial Distribution

Bernoulli trial

- **data reflects a finite (binary) classification**
- **an attribute is present or not, e.g., success or failure**
- **outcomes are independent**
- **probability of “success” is the same in each trial**

Binomial distribution

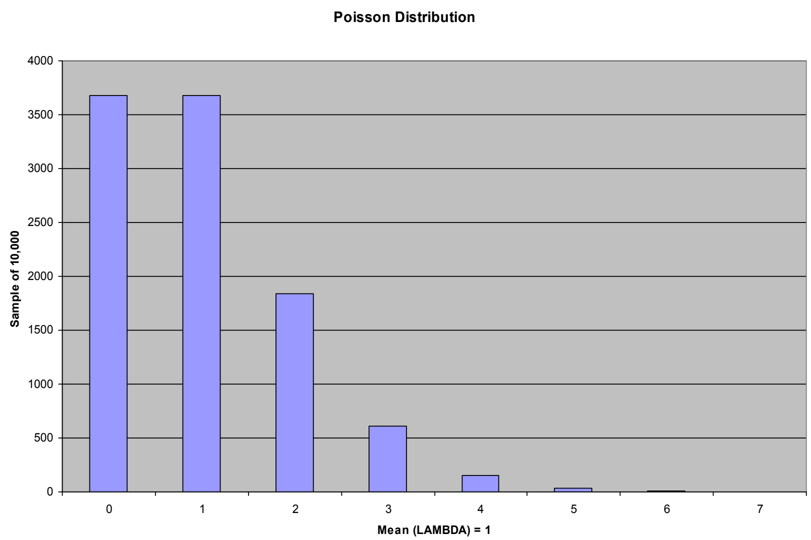
- **the number of successes in n Bernoulli trials**
- **y successes, p probability of success**
- $\mu = np$, $\sigma^2 = np(1-p)$

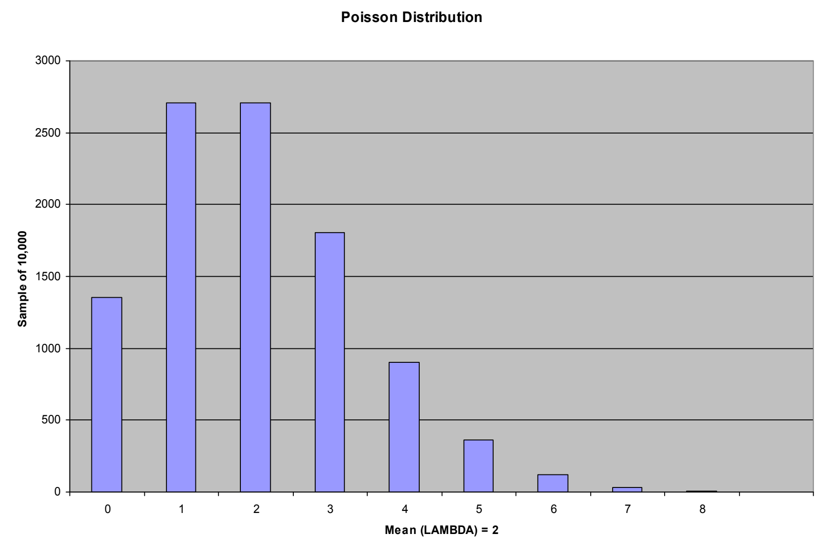
Poisson Distribution

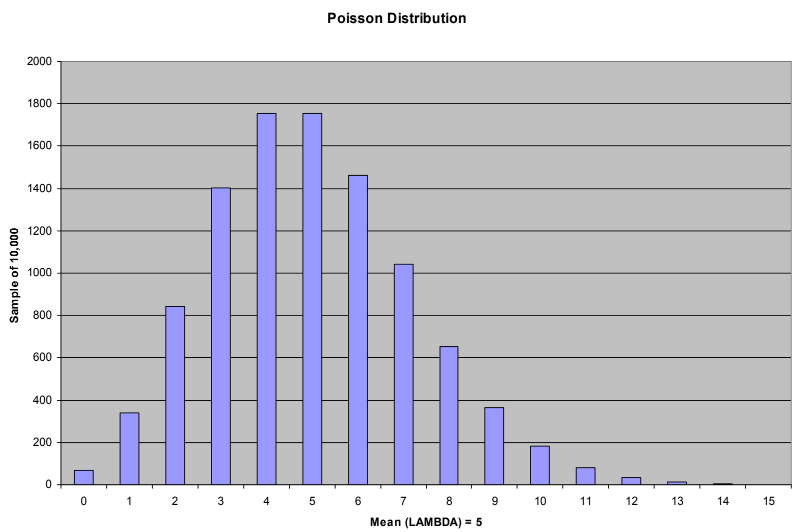
Poisson processes

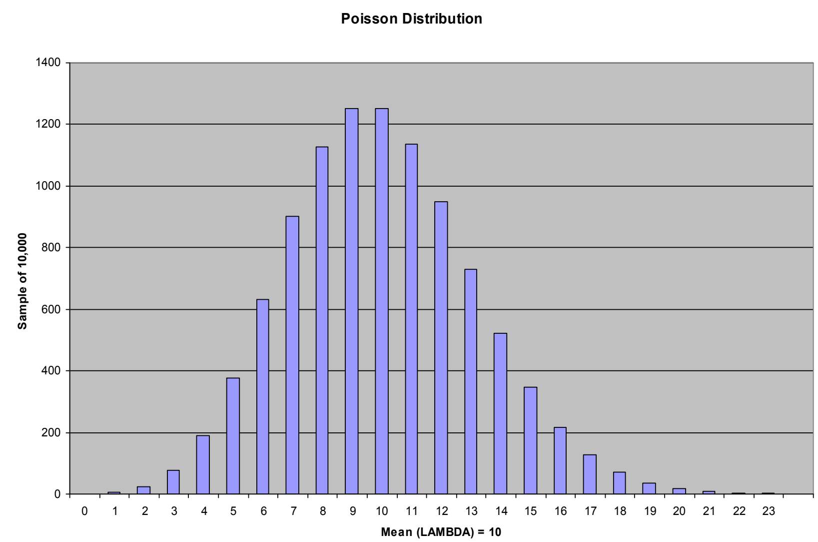
- **resulting data reflect a count of events**
- **the events are rare**
- **the events are independent of each other**
- **the likelihood of a given event is related to the number of opportunities there are for its occurrence**
- **limit of a binomial probability when $n \rightarrow \infty$ and $p \rightarrow 0$ such that $np = \lambda$**
- $\mu = \sigma^2 = \lambda$

What Is A Poisson Distribution?









Avoiding Overcontrol

By understanding variation, one understands that …

… overcontrolling (aka tweaking, tampering) processes increases variation

… reducing micro-management improves employee morale and performance

Statistical Thinking Summary

Measure a few critical things with sufficient attributes and at adequate granularity to do sophisticated analyses.

Good operational definitions are fundamental to quantitative management / management by fact.

Statistical thinking

- **all work is a series of interconnected processes**
- **all processes are variable**
- **understanding variation is the basis for management by fact and systematic improvement**

Rigorous Statistics

High maturity organizations use rigorous statistics, such as

- **control charts**
- **prediction intervals**
- **confidence intervals**
- **analysis of variance**
- **etc.**

when appropriate for answering various questions – providing business value.

The field of statistics cannot be adequately discussed in a handful of slides!

- **This slide deck is an “awareness” presentation.**

Workshop – Setting Expectations

"Higher maturity organizations typically have had success with statistical techniques. Understandably, they want to encourage lower maturity organizations to adopt these techniques as part of their process improvement activities. However, such enthusiasm must be tempered with the realities and how such movement is perceived by lower maturity level organizations. Applied too soon, to unreliable low maturity data, statistics can be time-consuming and counterproductive."

Seven Basic SPC Tools

Scatter diagrams

Run charts

Cause-and-effect diagrams

Histograms

Bar charts

Pareto charts

Control charts

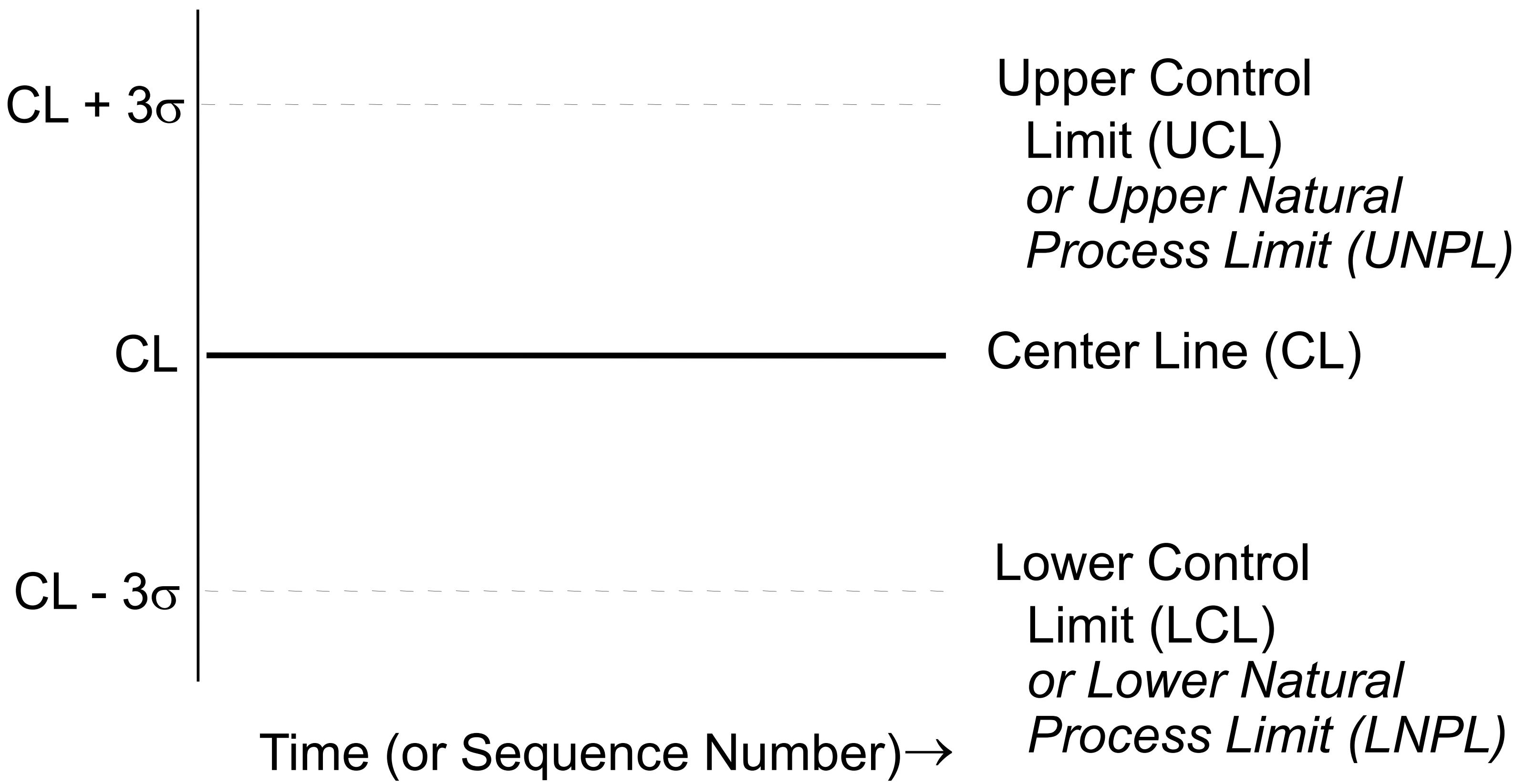
SPC Implies…

… control charts (aka process behavior charts)

… the use of a broad range of quantitative and statistical techniques

- **simple graphics**
- **brainstorming techniques**
- **causal analysis tools**
- **control charts**
- **classical statistical techniques such as regression analysis and ANOVA**

Control Chart Basics



Uses for Basic Tools

Chart trends over time

Examine relationships among measures

Explore cause-effect relationships

Prioritize issues

Understand the past …
Control the present …
Predict the future…

Decision Making Support

Use sophisticated analytical techniques when they are needed to understand the impact of a proposed change.

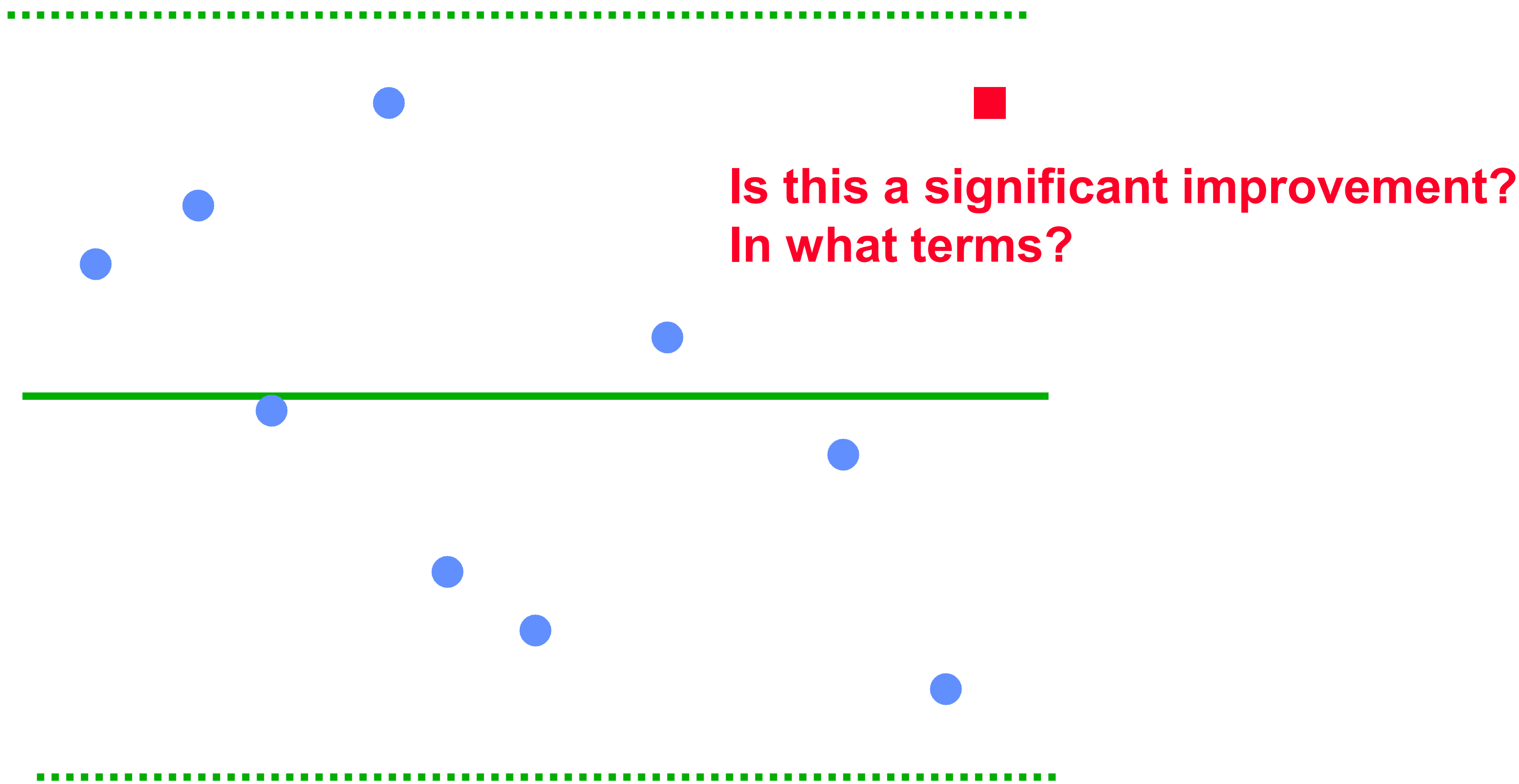
If the impact of a change is large and obvious, e.g., order of magnitude improvement, sophisticated analyses are unnecessary.

- **capture the evaluation consistently and using pre-defined criteria**

Remember: decisions are not just "by the numbers!"

- ***User feedback, e.g., in the form of user surveys after pilots, should be considered in deciding about deployment.***

Analyzing the Pilot Results



Long-Term Analysis – Trends

Even if piloted, some changes do not “scale up” well in full-scale deployment.

Tracking of performance after deployment is important for understanding the long-term impact of a change – and verifying its value.

- **data categorization is critical for dealing with confounding factors**

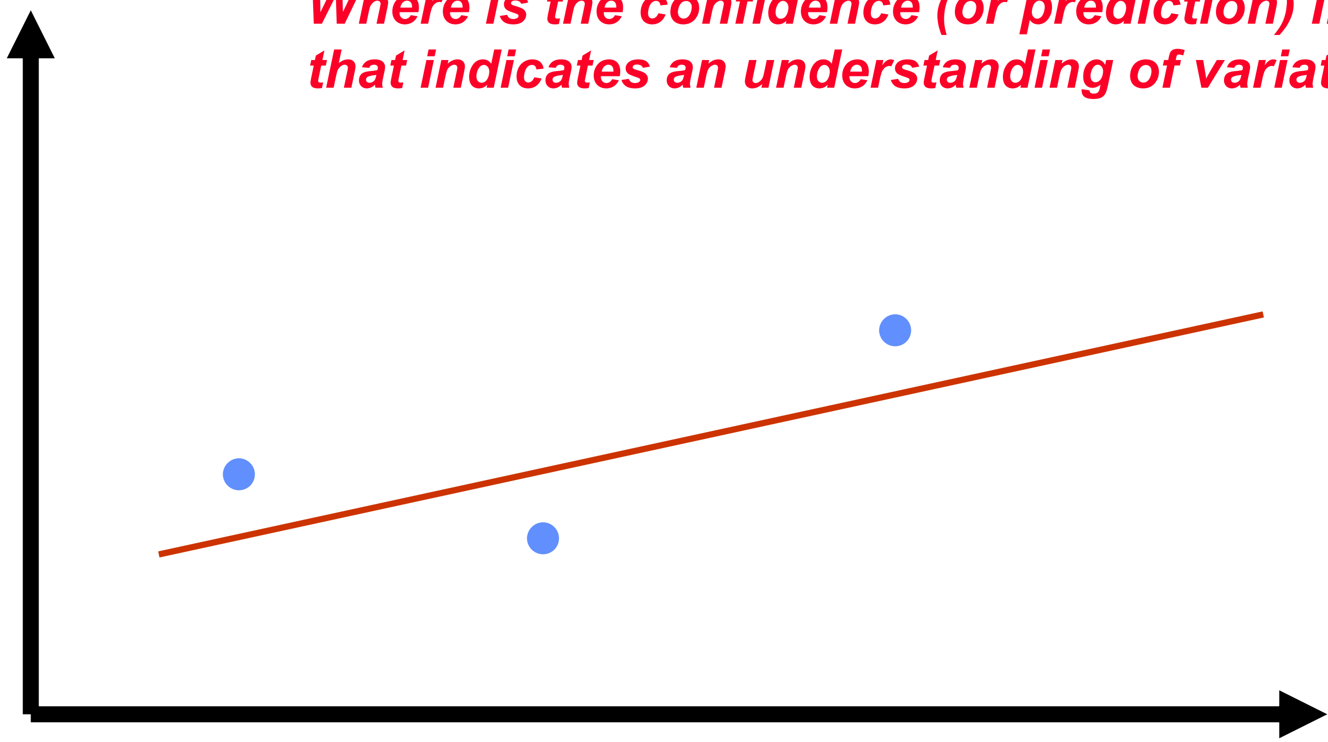
Process modeling, e.g., systems dynamics models, may be useful for understanding feedback and interaction effects.

- **use the subprocess data obtained at Level 4**

A Valid (Useful) Trend Chart?

Regression analysis is more than a trend line…

Where is the confidence (or prediction) interval that indicates an understanding of variation?



Valid Analyses

For understanding the impact of “small changes,” sophisticated statistical tools are essential.

Compound, incremental changes can have dramatic performance impact.

Confounding factors add “noise” to the data that frequently must be filtered out.

Process Variation

Shewhart's notion of dividing variation into two types:

- ***common cause*** **variation**
  - **variation in process performance due to normal or inherent interaction among process components (people, machines, material, environment, and methods)**
- ***assignable cause*** **variation**
  - **variation in process performance due to events that are not part of the normal process.**
  - **represents sudden or persistent abnormal changes to one or more of the process components**

Equation Form

total variation

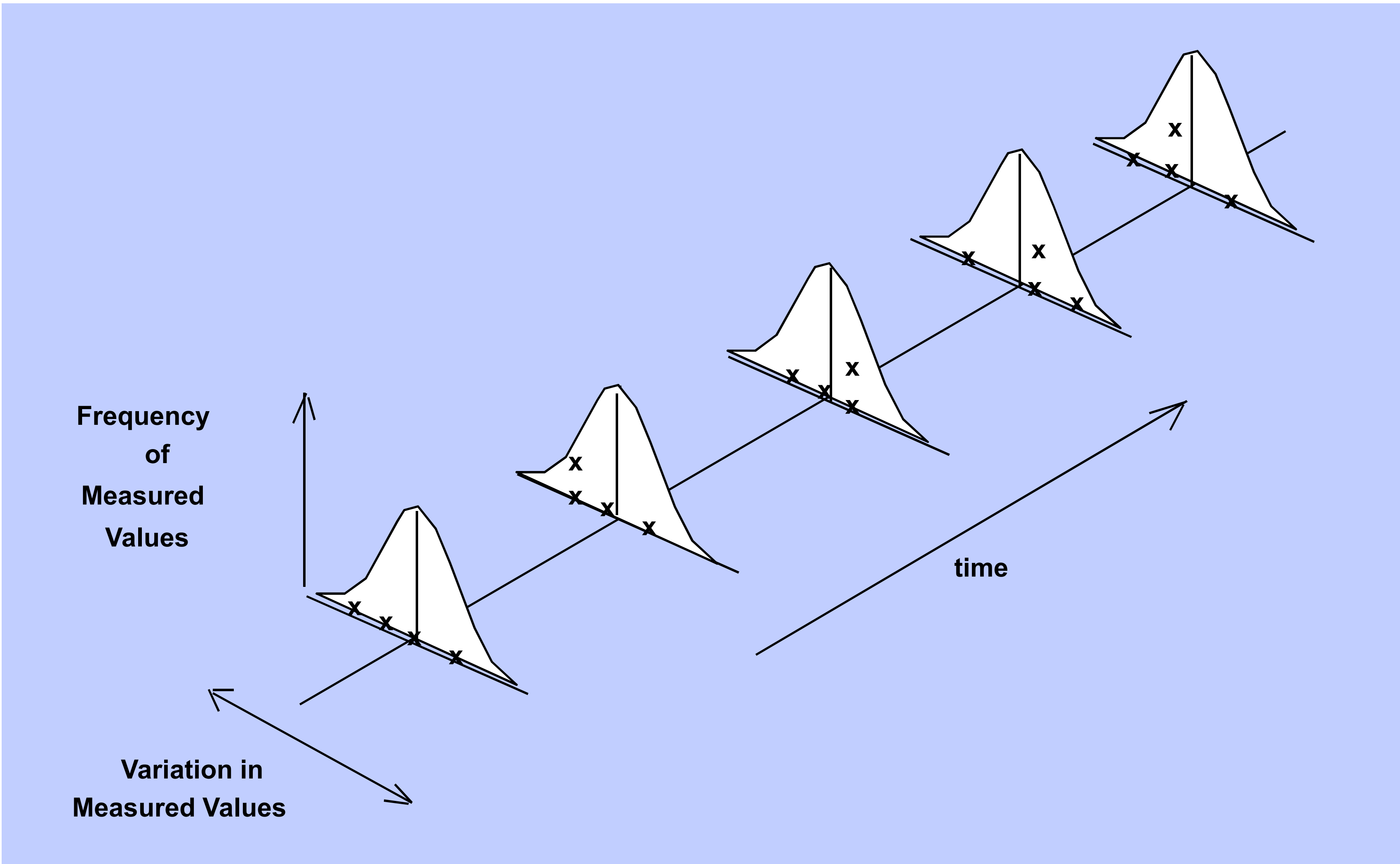
=

common cause variation

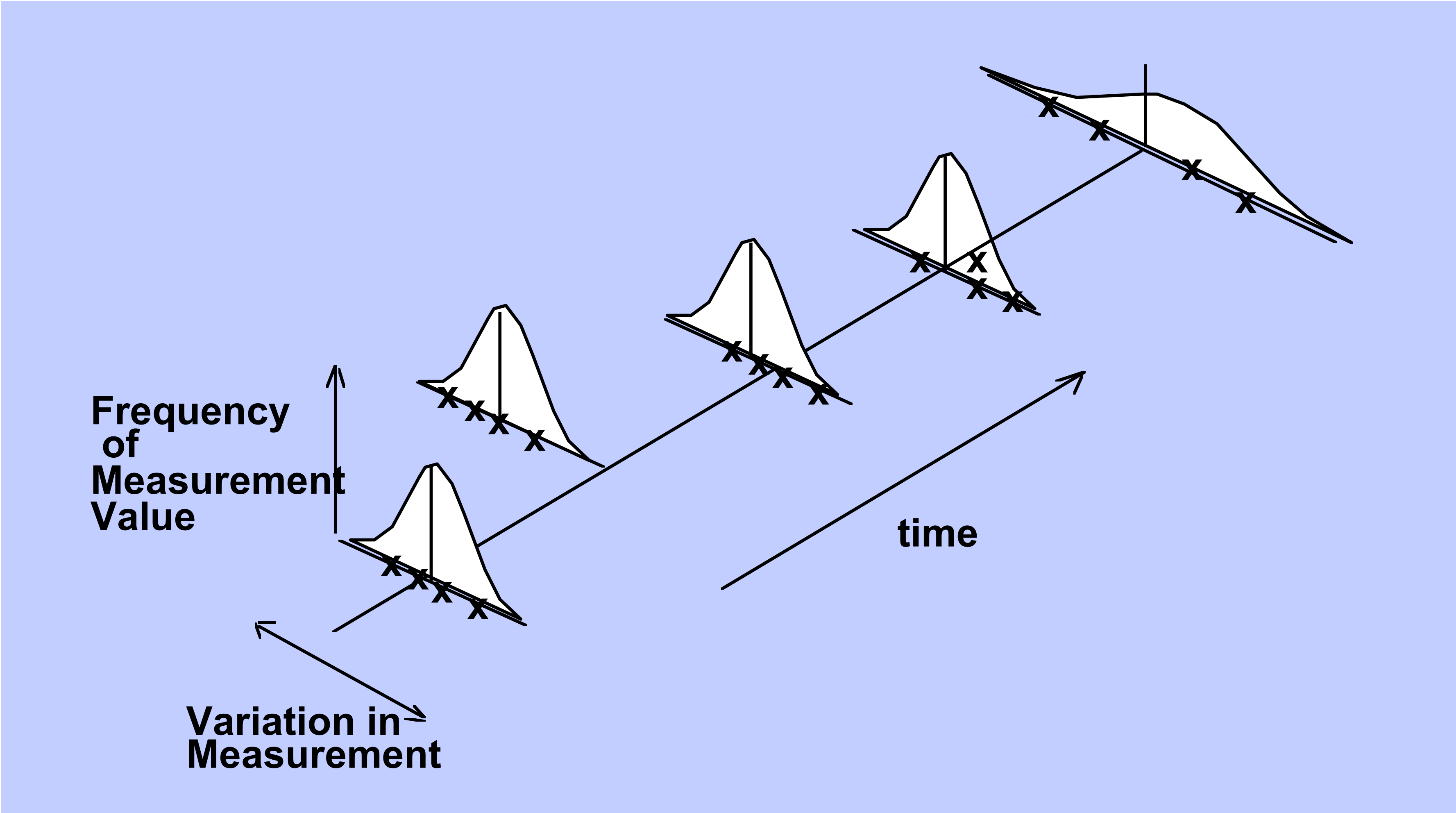
+

assignable cause variation

Example of a Process in Statistical Control



Example of an Out-of-Control Process



Listening to Voices

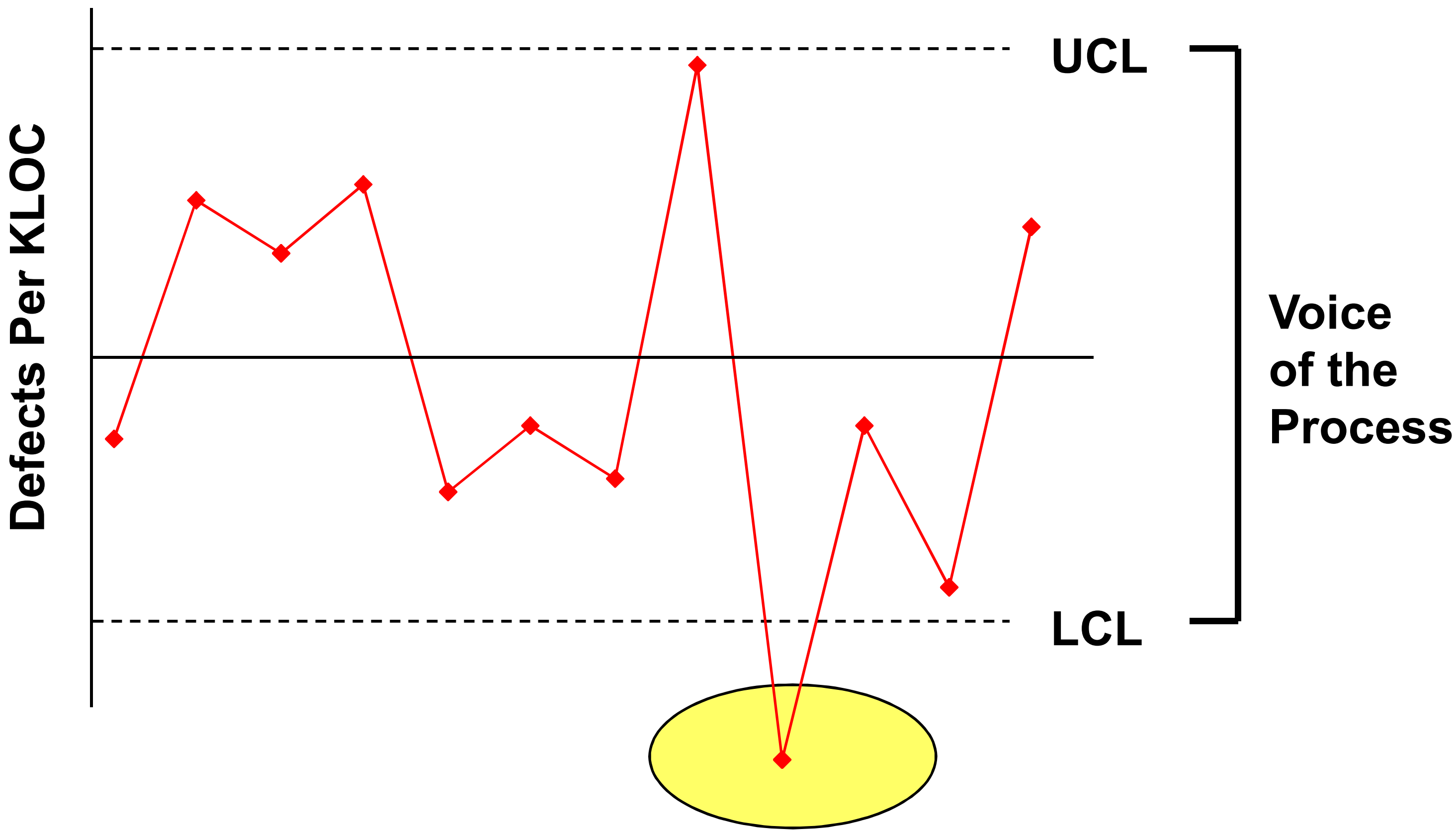
Voice of the process = the natural bounds of process performance

Voice of the customer = the goals established for the product and process performance

Capable process = { **stable process** **+** **product conformance** }

Process capability DOES NOT EQUATE TO capable process.

Control Chart – Signals



Detecting Signals

The simplest rule for detecting a signal (possible assignable cause): a point outside the 3-sigma control limits.

Many other sets of detection rules proposed.

- **usually run tests**
- **makes the control chart more sensitive to signals**
- **also leads to more false alarms**
- **decision on detection rules should be based on economic trade-offs**

Detection Rules

Test 1: A single point falls outside the 3-sigma control limits.

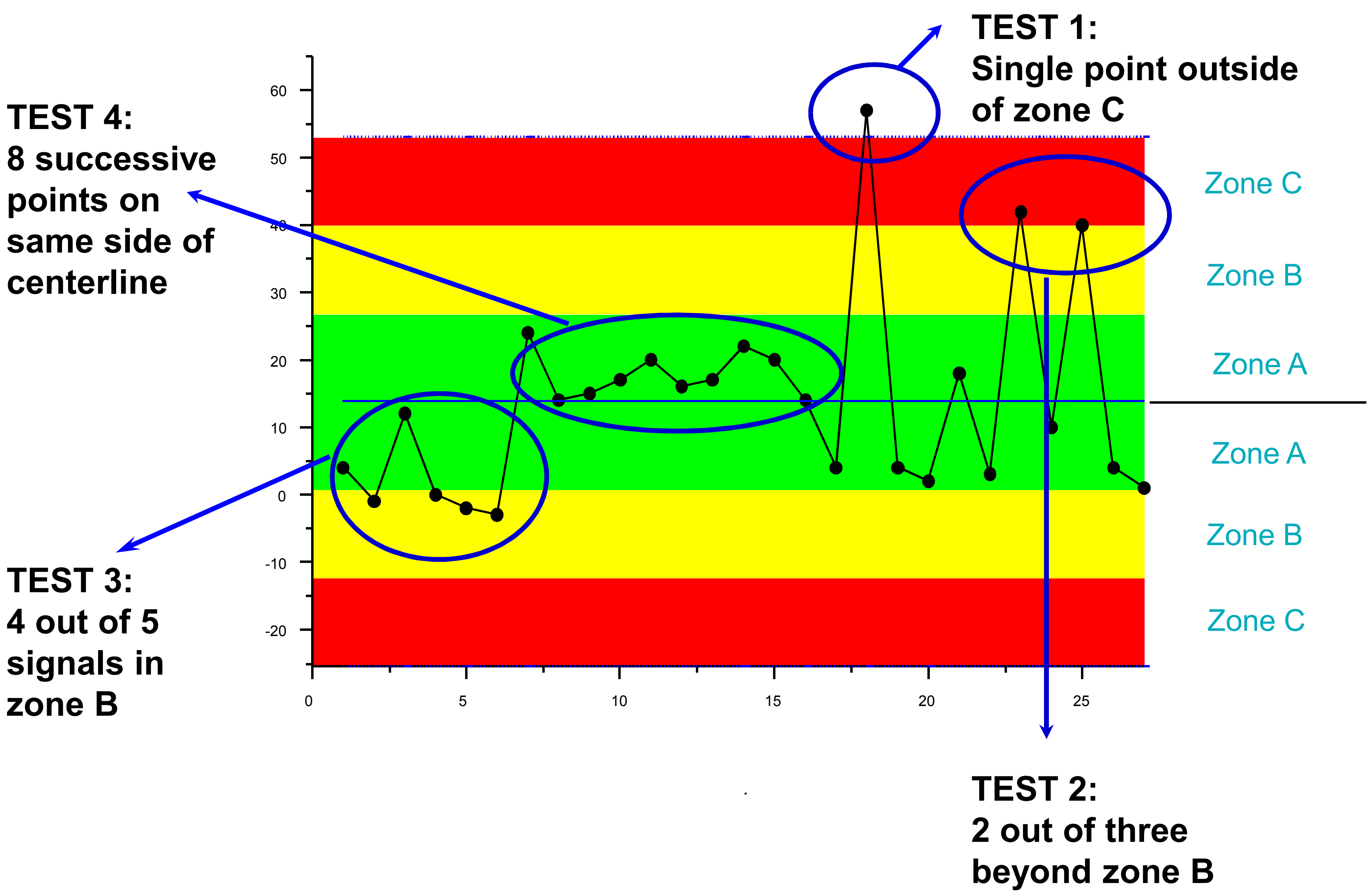
Test 2: At least two of three successive values fall on the same side of, and more than two sigma units away from, the center line.

Test 3: At least four out of five successive values fall on the same side of, and more than one sigma unit away from, the center line.

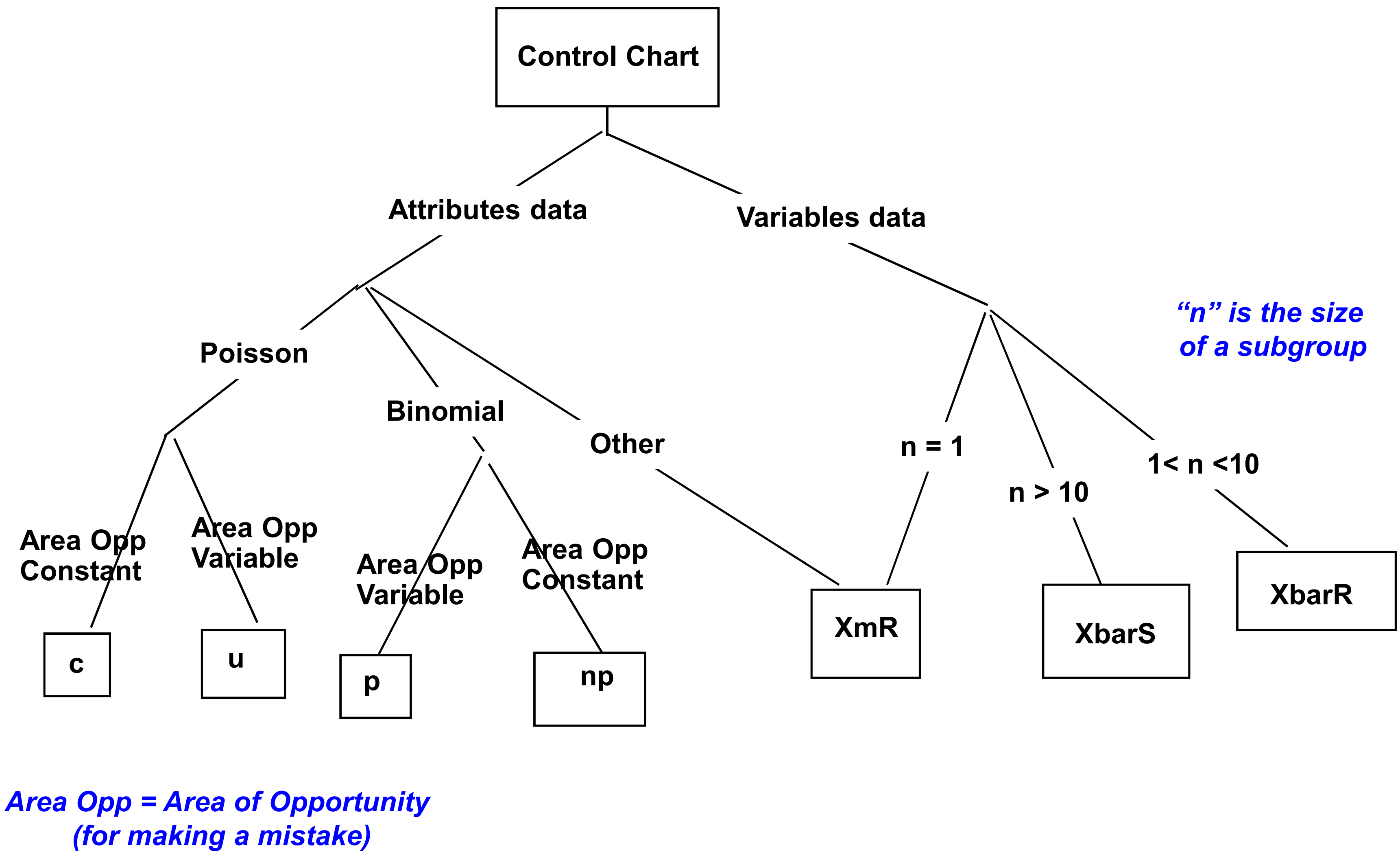
Test 4: At least eight successive values fall on the same side of the center line.

Source: Western Electric Handbook

Examples of Detection Rules



What Control Chart Should Be Used?



XmR Chart

Assumes population standard deviation (sigma) is constant across all observations

- **measuring a common-cause system!**

Does not assume a particular statistical distribution, e.g., Poisson or binomial data

Also known as individuals and moving range chart.

Weakness: "chunky data"

- **range with less than four values**
- **average count less than 1**

Control Limits for XmR Charts

Individuals Chart Limits

$$\text{Upper Natural Process Limit} = UNPL_X = \overline{X} + \frac{3\overline{mR}}{d_2} = \overline{X} + 2.660\overline{mR}$$

$$\text{Center line} = CL_x = \overline{X} = \frac{1}{k}\sum_{i=1}^{i=k} X_i$$

$$\text{Lower Natural Process Limit} = LNPL_x = \overline{X} - \frac{3\,\overline{mR}}{d_2} = \overline{X} - 2.660\,\overline{mR}$$

Moving Range Chart Limits

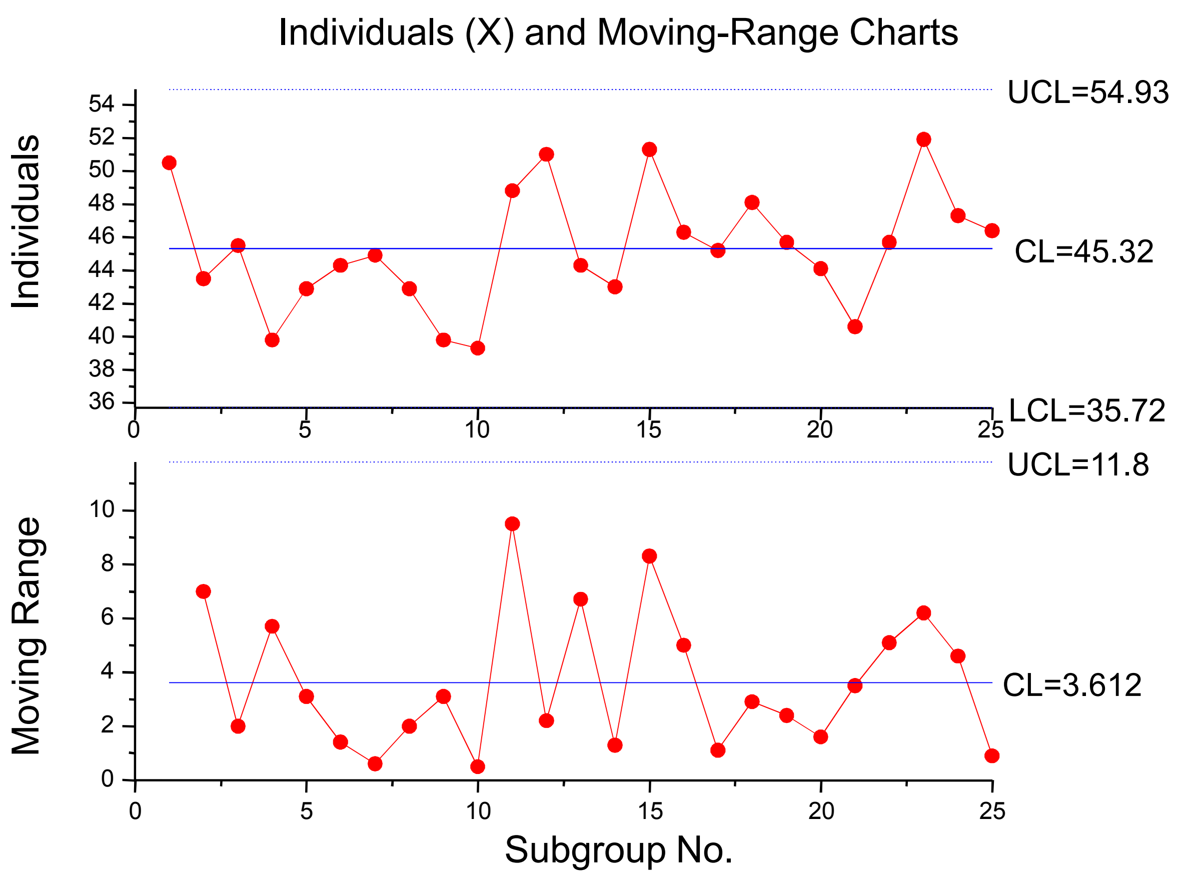
$$\text{Upper control Limit for moving range} = UCL_R = D_4\,\overline{mR} = 3.268\,\overline{mR}$$

$$\text{Center line or average moving range} = CL_R = \overline{mR}$$

$$\text{Lower control Limit for moving range} = LCL_R = D_3\,\overline{mR} = 0 \text{ for n = 2}$$

d_2 = 1.128 for n = 2

Example – XmR Control Chart



Myth of Normally Distributed Data

XbarR and XmR charts are <u>robust</u> in the presence of non-normality.

It is possible to construct pathological cases…

- **pathological in sense of economics**
- **Tchebycheff's theorem: at most 11% of data outside 3σ limits**
- **Wheeler's Empirical Rule #3: given a homogenous set of data, approximately 99-100% of the data will be within 3σ of average**

Some types of control chart take advantage of knowledge of distribution, e.g., u-chart for Poisson.

Myth of Too Much Variability

If there is a great deal of variability in the data, then the control limits will be wide

- **may indicate highly aggregated data**

High variability has consequences

- **if the limits are wide, predictability is poor**
- **if highly variable performance is "unacceptable," then the process will have to be changed**

Performance of people varies -- a lot, especially in an ad hoc environment…

… but ignoring reality will not change it

Myth of Insufficient Data

Is there only one data point per project or phase?

Control charts should be applied to individual processes, e.g., code inspections.

- **30K SLOC program**
- **100 inspections (data points) at 300 SLOC per inspection**
- **preliminary control limits can be established with 4 or 5 data points**
- **20 data points is usually sufficient for reasonable control limits (40 is desirable)**

Myth of Living on Internet Time

Reality check: software processes and products change "rapidly."

Do processes change significantly during a phase? Implies instability…

Are "ripple effects" significant?

- **of changes on downstream activities?**
  - **changes in inputs will change performance**
- **of (planned) changes on upstream activities?**
  - **possibility of feedback loops or anticipation**
- **unless bootstrapping with data from other projects (small run/short run environment), why should ripple effects be a concern?**

Multiple, Overlapping Processes

Mixing and stratification of data from multiple processes that are lumped together?

Begin with "informally stabilizing" the process, refining operational definitions

Understand the data before trying to control the process!

The first major challenge is one of good operational definitions – just like Shewhart said.

Incorrect Statistical Techniques

Most common error: using standard deviation as the estimator for sigma

- **several ways to estimate sigma**
- **for an unstable process, standard deviation combines variance of assignable causes and common cause system - limits are too wide**
- **control chart formulas use bias correction factors**

The second major problem is understanding the last 70 years of (basic) SPC work – which even statisticians don't always do well!

Management Training

Need to train managers on how to use measurement and statistical control

- **effectively to perform their roles**
- **in ways that will not cause dysfunction**

Managers do not "control" the engineering process - engineers do.

- **managers allocate resources, resolve conflicts, and enable engineers to function**

The third major problem is avoiding dysfunctional behavior.

The Bottom Line on Statistics

Real-world software organizations, such as Onboard Space Shuttle and Motorola India, are using control charts and other rigorous statistical techniques and reporting <u>business value</u>.

SPC has many tools – use the right tool to answer the <u>business</u> questions you have.

Questions and Answers