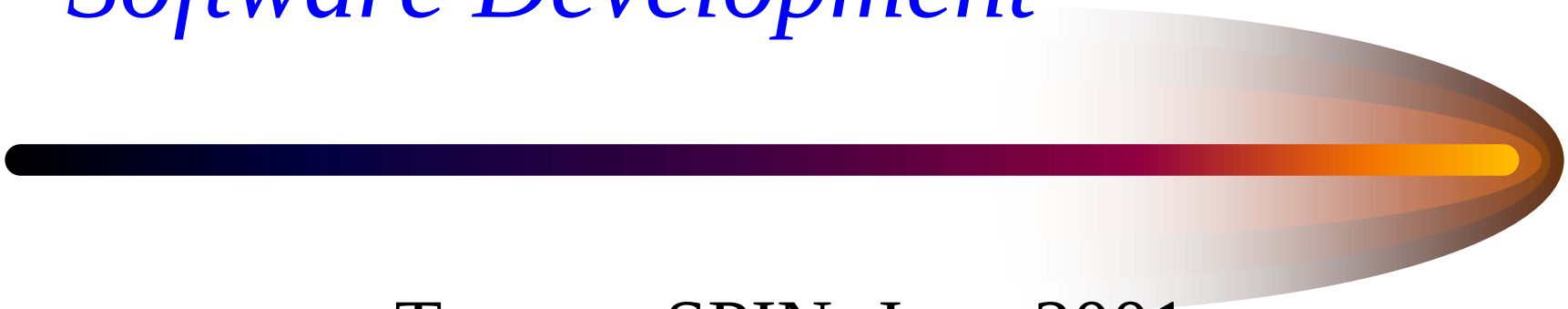
Uses and Abuses of Metrics in Software Development



Toronto SPIN, June 2001

Pat Cross

London St. Consulting

Everyone loves metrics

- CMM Software has recently moved the requirement for gathering software metrics from Level 3 to Level 2

- ISO 9000-2000 now requires ‘monitoring and measurement’ of processes and products

Everyone loves metrics

The questions are -

What do they think it will buy us?

and

Why aren’t we all doing it anyway?

Simple-Minded Answers

- They think using metrics will help us increase quality

- Many of us avoid metrics because we think they are expensive, time-consuming and we can't prove to the company that they have a positive payback

Manufacturing Metrics

- Over the years, Manufacturing learned to describe defects so they could be measured
- They used key measurements and SPC to control the defects
- Later, they used other analyses to point out how they could make the product better

This took almost a century to develop!

Uses of Metrics

Typically, measurements are used to:

- Fact-find
- Predict characteristics
- Prove improvements
- Standardize products
- Predict improvements

Examples of the Uses of Metrics

Fact-find - determine the acceptability of your product to the user

Predict characteristics - write a project schedule

Prove improvements - compare before and after

Standardize products - force certain characteristics to stay within a known range (tolerance range)

Predict improvements - forecast which aspects of process or product will have the largest effect on quality for a specific project

Common Statistical Methods

Fact-finding - large deviation from expected, Pareto analysis, central tendency

Predict characteristics - central tendency, trend graphs

Prove improvements - proof of hypothesis

Standardize products - SPC (control charts)

Predict improvements - Design of Experiments, Taguchi Analysis

Look before you leap

- Each of the different objectives will have different metrics to help achieve it
- Before you spend one cent on writing down a metric, DECIDE WHAT YOUR GOAL IS
- Metrics gathering and analyzing is expensive (and boring). Notwithstanding CMM and ISO 9000, don't do it unless you mean to use them

Goal-Question-Metric

- A classic way of deciding what metric should be measured is GQM
  - decide a goal
  - pose one or more questions about meeting the goal
  - provide one or more answers to the questions
  - see what can be measured in the answers
  - pre-plan how you will use the results

Example of GQM

Suppose you are planning to introduce a new process,

The goal is “check if it is an improvement, or just a change”

The question is “what would an improvement look like?”

That should lead to a metric which should be measured before and after the change.

Measurement Theories

- There are many well articulated measurement theories, first developed for hardware
  - Several different types of measurement exist
  - Each of them have strengths and weaknesses
  - Each of them have different statistical methodologies
- Don’t use a metric whose limitations you do not understand

Types of Measurement

One set of major types is:

- Discrete, eg simple counting (integral), called attribute by manufacturing
- Continuous eg a length, called variable by manufacturing
- Non-parametric (numbers aren't used) eg how sour is this? Very sour, quite sour, not sour

Types of Measurement (2)

Another set of major types is:

- Direct

  the parameter you are interested in can be directly measured eg the time to complete preparation for an inspection

- Indirect

  the parameter you are interested in cannot be measured directly- you must use an indicator eg McCabe's measure for complexity

Trusting measurements

Sometimes decisions, predictions and justifications are made on the basis of one measurement, but mainly people would prefer repeated measurements to increase their confidence. They take a ‘sample’ of the population

The larger the sample with a close cluster of measurements, the more they trust their validity

Condensing measurements

Example

Inspection team members are asked to measure their preparation time to the nearest 15 minutes

From the following numbers, how would you budget inspection preparation on a project?

Example of Mean, Mode and Median

Time reported	# of people reporting
30	3
60	54
75	15
150	2
375	8

The mean is 90 minutes

Precision of measurements

- In the last example, the precision of the time measurements was limited to 15 minutes
- If we had not put that restriction on, the precision would have been limited by the clock or by the person reading the clock
- Variable measures have the precision of the measuring system
- Discrete measures have a precision of the counting step

Confidence Levels

In the example, we had 82 data points and we used the mode or mean of that group to approximate the mode or mean of all inspection preparations ever undertaken or to be undertaken in that company.

The bigger the sample, the more confident we are that the sample mean is close to the population mean. Statistical methods allow you to calculate the confidence limits. The larger the confidence limit, the less confidence you have in the mean.

Comparing results

If you are doing a comparison (say before and after) to decide if a metric has changed, you have to

- be careful you have preserved the right precision, even though you have made manipulations to the data
- be careful you have taken the confidence limits into consideration. If the sample sizes are large and the difference between the group data is large, you can say the data are different - other wise, use statistical tests of significance to prove it

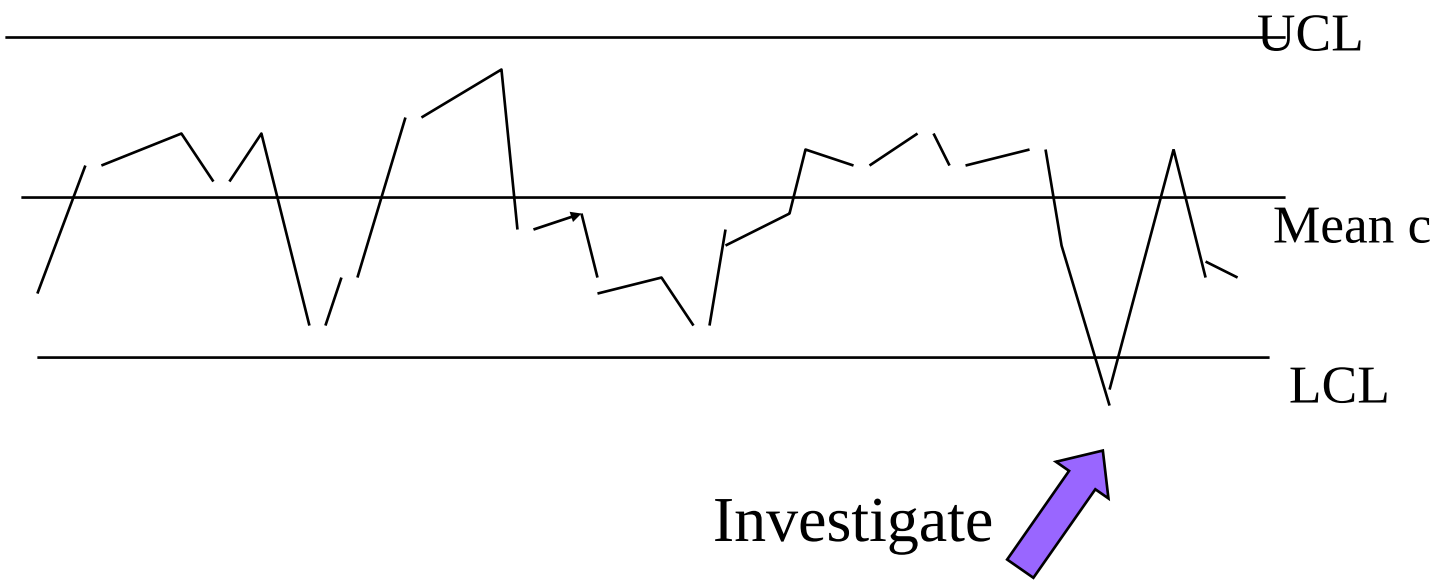
Controlling characteristics

- The standard version of Statistical Process Control (SPC) assumes random (uncontrollable) Normal variation of every characteristic about a mean over a well defined range (between the control limits)
- Changes outside these limits are likely to be unusual and controllable factors
- The object to to keep key characteristics within the limits - Act only if the characteristic is out of control

Example of SPC -c Chart

- For at least 20 units, count the number of defects per 100 LOC (c) found in unit testing.
- Check for Normality. If it more or less normal, plot c against unit ID
- Draw control limits (mean c $\pm$ 3 times the square root of mean c)
- sample the unit test results and continue plotting
- investigate only when out of control

c Chart for Defects in Unit Tests



Abuses of Metrics (1)

1. Gather data without a goal in mind. Hope to find something to use it on
2. Gather one type of data (eg discrete) and treat it as if it were another type (eg continuous).
3. Don’t consider precision
4. Forget that an indicator is not a characteristic

Abuses of Metrics (2)

3. Use statistical methods without understanding the theory of the methods
   - use statistical methods for variables when you are gathering non-parametric data
   - don't use 'tests of significance' when comparing two groups of data
   - don't test your distributions for Normality before using manufacturing-type stats
   - don't use text book control limits

Abuse Examples

1 A telecom company in Toronto which insisted defects from inspections should be characterized in three ways

- importance level, high, medium or low
- defect type, data incorrect, missing or superfluous
- presumed origin, phase of project

This made the process cumbersome; but they only used the data for immediate remedial fixes

Abuse Examples

2. A research company in Toronto used a checklist of 15 questions on 17 software groups to form an opinion on their maturity. The answers were Yes/No (ie data was discrete). They were translated into percentages and presented as continuous. The companies were sorted into a group of 8 and another of 9. The means of the groups were quoted as 48.53% and 52.48%. It was suggested Group A could learn something from Group B

Abuse Examples

The precision was 1/15. The data should not have been presented as percentages and if it was it should have been presented as a range 45 to 51%, 49 to 56%. Nothing could justify the number 48.53%. Clearly the precision ranges would overlap.

In addition the confidence limits at the 95% confidence level were more than 3% . Even if the precision had been better, the sample size would not support the assertion that these two groups were different.

Abuse Examples

3. A bank in Toronto which drew control charts for schedule times
   - didn’t check the distribution was Normal
   - used raw data rather than sample means
   - didn’t use a range chart
   - used 1 sigma instead of ‘amended 3 sigma’ for the control limits
   - didn’t use them for control, but fact finding

So where do we go from here?

My opinions are:

- Metrics are essential to move beyond the basics in achieving quality
- Metrics are expensive and labour-intensive
- Without a knowledge of simple measurement theories and appropriate statistics, you will waste money and not achieve your objectives

Basic Steps Forward

- Take time to chose goals which align with the company and start simple
- Take time to learn about the metrics you want to use, their strengths and weakness for you
- Take time to learn about the assumptions and limitations of any analysis method you use