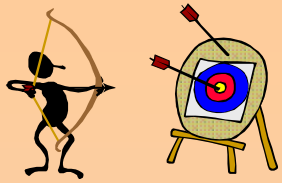
SITARA
Excellence Combining Expertise™

Examples of Application of Quantitative and Statistical Methods in Software Engineering

Raghav S Nandyal
Chief Executive Officer
raghav_nandyal@SITARATECH.com

SITARA
Excellence Combining Expertise™



LEARNER OBJECTIVES

- How does one apply quantitative and statistical techniques to software engineering measures?
- What are the relevant software engineering practices where quantitative and statistical analysis makes sense?
- Are there alternative quantitative techniques which can be explored for arriving at the most valid and beneficial analysis?

Some Perspectives to the Problem …

(1 of 4)

Blind reuse of quantitative and statistical methods that hold true in a manufacturing context, to the software development context is fraught with errors (need common sense logic for interpretation)

- Use of **random sampling** for product inspections
- Use of **statistical process control tools and Design of Experiments** for reduction of variability
- **FMEA and root-cause analysis**: *in software development, change is constant*; impact analysis is complex with unintended fan-out
- Emphasis of software measurement and analysis until recently had no connection to understanding business performance

Some Perspectives to the Problem …

(2 of 4)

Incorrect understanding that: a high-maturity process is different from a lower-level process; or that –

- Institutionalizing high-maturity practices at levels 4 and 5 requires a different approach to be used by projects in the organization
  - Something ***new and sophisticated*** happens
  - A ***new type of process design*** based on sub-process control and Quantitative Process Performance Objectives is required

- This mindset is probably due to Six Sigma's emphasis on Design For Six Sigma (DFSS)
- No such thing *actually* happens!
- The same old process is executed with an eye for detail, while being sensitive to variation and learning

Some Perspectives to the Problem …

(3 of 4)

Current emphasis of process prediction seems restrictive; it seems to suggest that a regression equation (modeling an outcome based on factors) is the **only** way to predict an outcome variable

- Prediction does not *only* mean that one is able to tell exactly where the next point on a specific metric will lie
- **It is also about predicting the values of other related measures knowing what one knows about a particular metric**
- For example: PCE at 100% in the code, unit testing and integration phase, means FPY in system testing should be 100%.
- Or, consider the example of Defect Density

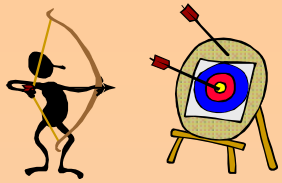
Some Perspectives to the Problem …

(4 of 4)

Wrong belief that high-maturity process **under statistical control** is already operating at its *best* and therefore, it is *impossible* to improve it further

- **"If you need a better performance, then it would require you to change the process or me, the person executing this process"**
- Not quite true!
- We will review examples in this presentation where even when a process is under statistical control, it can still be improved without changing the process or the people executing the process
  - Productivity tools and automation
  - Extensive dependence on reuse

SITARA
Excellence Combining Expertise™



AGENDA

- Establish the Context
- Do we need this measure?
- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- SITARA TQI for TQA^{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A

Establish the Context

(1 of 3)

Typical questions that confronted organizations from where these examples were sourced …

- "Our customers are satisfied with a level 2. We are a small Company with happy customers. Why do we need to focus on high maturity practices?"
- "Would a focus on building a high-maturity process not shift our focus away from developing software, to statistics?"
- "Why should we invest in high-maturity measurement and analysis when it is limited in context and applicability?"

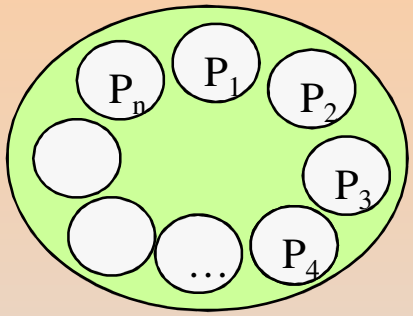
But, they answered questions based on a well structured high-maturity probe (SITARA TQI for TQA^{sm}), and have reaped rich dividends

Establish the Context

(2 of 3)

Application of Quantitative and Statistical Methods in Software Engineering is required primarily for 3 reasons –

- Summarizing Data,
- Comparing Data and,
- Interrelating Data



Myth 1: The competencies required to do them are complex and difficult to master!

- Not quite!

Myth 2: To be considered a high-maturity process, we need to collect data on everything that moves!

- Measuring everything that moves would be akin to counting the quills on a Porcupine [DeMarco 1995]

Establish the Context

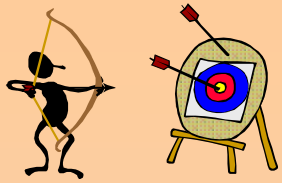
(3 of 3)

Based on the business objectives, quantitative and statistical methods in software engineering can help to improve organizational understanding of –

- both, process factors and people factors,
- that contribute to minimizing process variation, and improving organizational learning,
- to arrive at meaningful conclusions on the process capability

Useful to note that:

- Quantification and statistical methods are "typically" applied on the specific practices of the CMMI
- More specifically, on the engineering and project management categories; but, …
- **What are the right measures?**



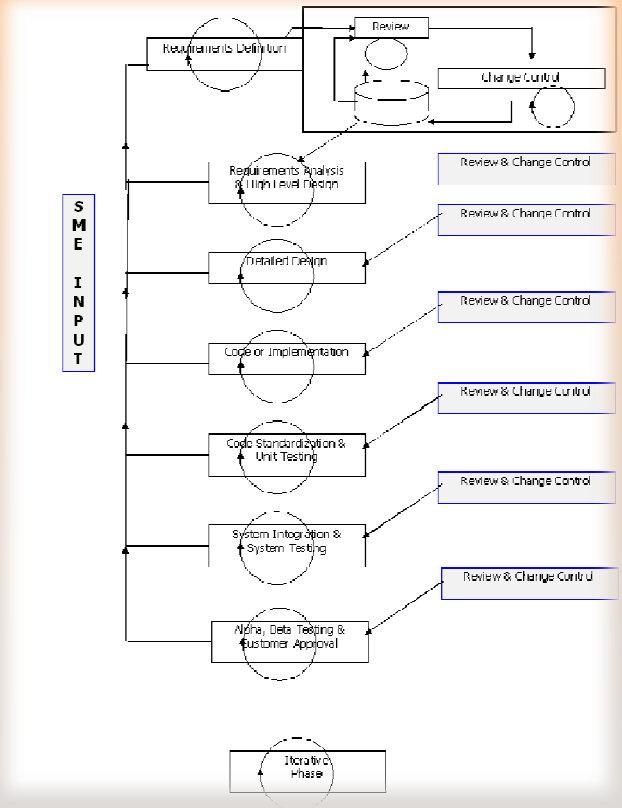
Do we need this measure?

- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- SITARA TQI for TQA^{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A

SITARA
Excellence Combining Expertise™

Do we need this measure?

(1 of 3)



Basic Metric Data
Code size
Effort
Schedule
Problem Counts
Staff
Requirements
Milestones

Derived Measures	
1.	Requirements volatility
2.	Staffing
3.	Cumulative milestones
4.	Schedule variance
5.	Estimation accuracy
6.	Productivity
7.	Code distribution
8.	Effort distribution
9.	Fault distribution
10.	In-process faults
11.	Phase containment effectiveness
12.	Cost of quality / poor quality
13.	COQ - Appraisal effort
14.	Prevention effort
15.	Internal failure effort
16.	External failure effort
17.	Q-Data
18.	Inspection rate
19.	Preparation rate
20.	Error density
21.	Inspection fault type distribution
22.	Age of open problems
23.	Age of closed problems
24.	Phase Screening Effectiveness

You need data to see what is inside the process; a lot of data!

SITARA
Excellence Combining Expertise™

Do we need this measure?

(2 of 3)

Related to this question are the following questions–

- Why do we need this measure?
- What do we do with the measure once we have it?
- What type of a "process input" does this measure cater to?
  - **Controllable factor**
  - **Standard Operating Procedure**
  - **Noise**

The most typical management concerns in a software project are –

- How much longer will it take to deliver the product?
- After the product is delivered, what are the typical issues that we are likely to encounter?

Do we need this measure?

(3 of 3)

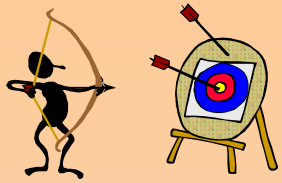
To answer these two questions meaningfully –

- A thorough design and understanding of the process is required because,
- it is in the lifecycle phases that time is consumed and errors or defects injected and detected

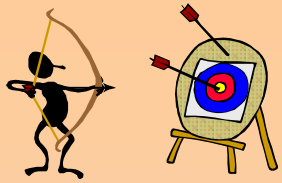
Key point: Measurement should be an integral part of the development lifecycle (e.g., sample mean, standard deviation, sample size)

Key point: Data must be collected and analyzed in real-time using simple techniques for it to be useful

Key point: There is little meaning in using software data as in a typical post-mortem – after the fact



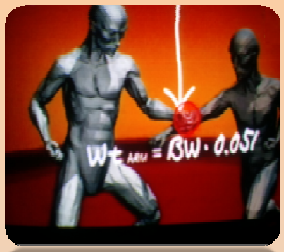
- Do we need this measure?
- **How do we decide the most appropriate measure?**
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- SITARA TQI for TQA^{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A



How do we decide the most appropriate measure?

(1 of 4)

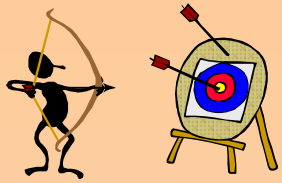
This question relates to answering –



- What aspects of process performance do we like to understand ?
- Fundamentally: it requires an accurate collection of effort or time, cost, size, error and defect counts – *both actual and estimated* – **of critical lifecycle phases**
- Establish derived measures using some combination of these inputs for –
  - **Summarizing**
  - **Comparing**
  - **Interrelating**

Using which, a **general** from the **particular** can be derived or established

SITARA
Excellence Combining Expertise™



How do we decide the most appropriate measure?

(2 of 4)

What do we mean by – establishing a "general from the particular"?

- Deterministic Model: **F = Mass * Acceleration**
  - Holds true, if **the physics** behind this equation has not changed
- Probabilistic Model: **COQ = 0.147 + 1.04*Appraisal Cost + 0.679*Prevention Cost + 0.998*Failure Cost** + *e*
  - Holds true, if **the probability and statistics** behind this equation has not changed
  - Conditional prediction model
  - Becomes the cookie-cutter with which COQ for projects of a similar type can be estimated

How accurate is the resulting probabilistic model?

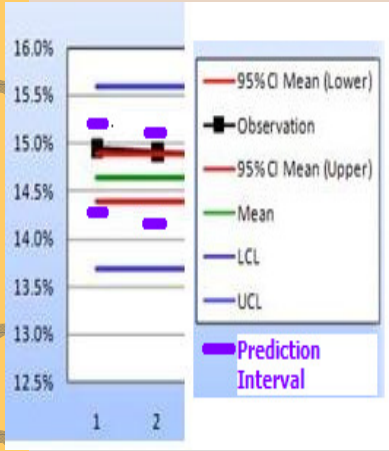
How do we decide the most appropriate measure?

(3 of 4)

Key point: The resulting probabilistic model is always "approximate"

- Depends at least on the number of estimators used, and accuracy of the estimators

Key point: Since being exact is impossible and, being 'approximate' is not good enough, we use prediction and confidence intervals



- While summarizing software data from a sample for say, effort variation and post-release defects density, we use the average value–**single value estimate**–
- resulting from a number of individual values
- **Prediction and Confidence intervals** puts this single value in right perspective since such precision is not guaranteed

How do we decide the most appropriate measure?

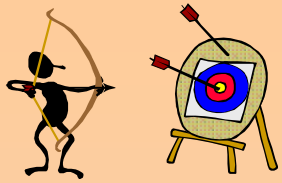
(4 of 4)

Key point: A regression model is a *conditional prediction model* …

- since the dependent variable COQ is accurate only if we can accurately establish Appraisal Cost, Prevention Cost and Failure Cost,
- these regression or independent variables themselves assume a range of values

Can we explore an example of a regression model?

What does a regression model serve?



- Do we need this measure?
- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- SITARA TQI for TQA^{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A

Example of a Regression Model

(1 of 1)

Regression Analysis: Org COPQ using Reqt. COPQ, Design COPQ, CUT COPQ, Test COPQ

The regression equation is
Org COPQ = 0.147 + 1.04 Reqt. COPQ + 0.679 Design COPQ + 0.998 CUT COPQ + 1.00 Test COPQ

Predictor	Coef	SE Coef	T	P
Constant	0.1470	0.2227	0.66	0.538
Reqt. COPQ	1.03542	0.04228	24.49	0.000
Design COPQ	0.67922	0.03940	17.24	0.000
CUT COPQ	0.998249	0.009767	102.20	0.000
Test COPQ	1.00440	0.00756	132.90	0.000

S = 0.00679366 R-Sq = 67.0% R-Sq (adj.) = 59.3%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	4	94.371	23.593	511176.82	0.000
Residual Error	5	0.000	0.000		
Total	9	94.371			

Source	DF	Seq SS
Reqt. COPQ	1	43.804
Design COPQ	1	49.715
CUT COPQ	1	0.036
Test COPQ	1	0.815

Interrelating Data

Example of Summarizing Data

(1 of 1)

Metric	Unit of measure	Q1 YYYY performance		Q2 YYYY performance					
		Average	Std. Dev.	Sample Size	Min.	Average	Max.	Std. Dev.	95% CI Mean
COPQ	%	10.66*	0.58	30	14.26	14.64	15.12	0.31	14.4-14.88
COQ Appraisal	%	16.88	0.97	30	18.98	22.27*	22.53	1.35	20.06-22.44*

**Performance is based on median data since the population did not follow normal distribution.*

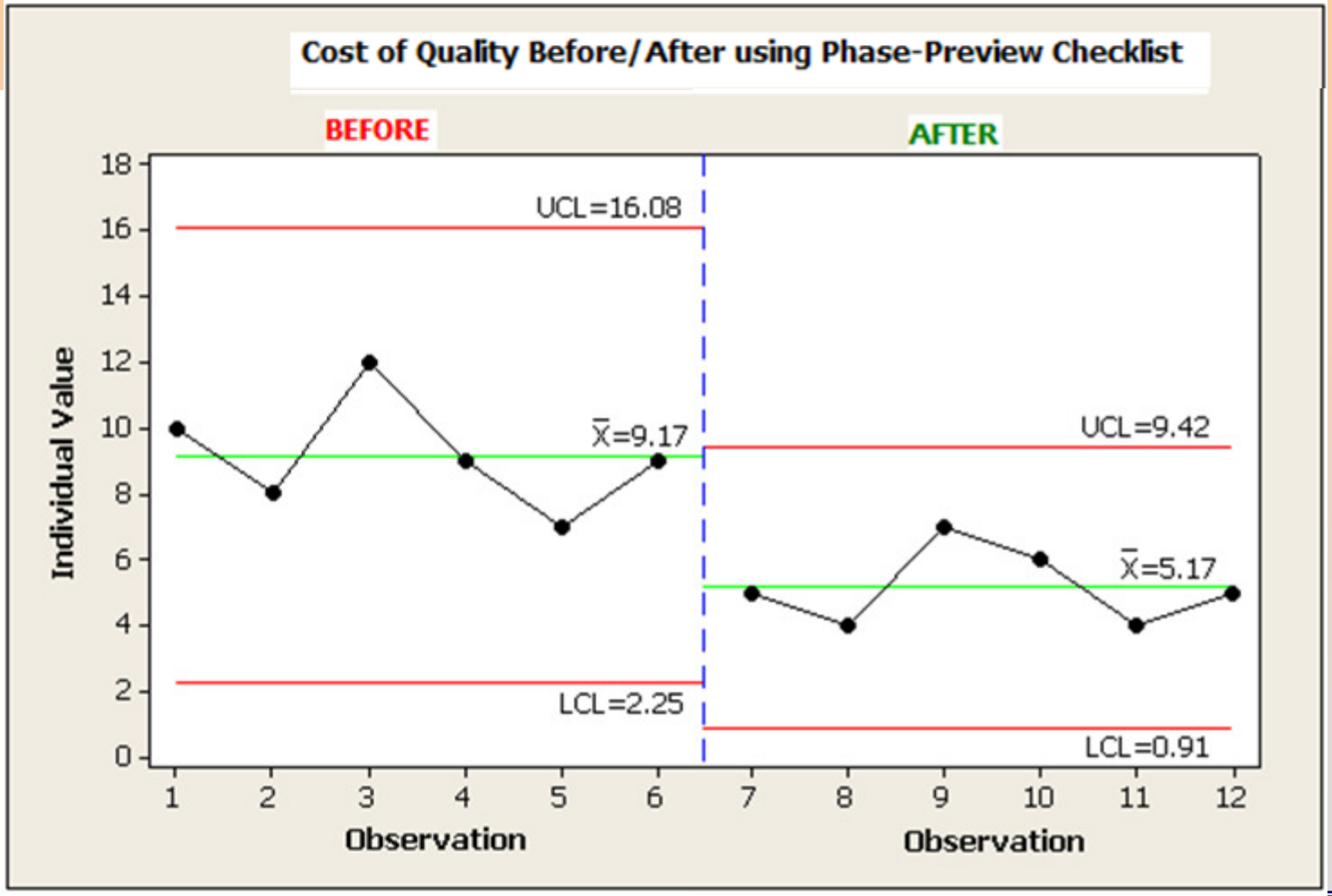
One Way

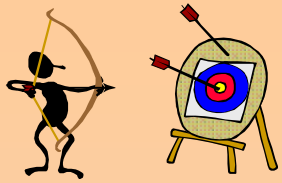
A More Useful Way

And, there could be others…

Example of Comparing Data

(1 of 1)





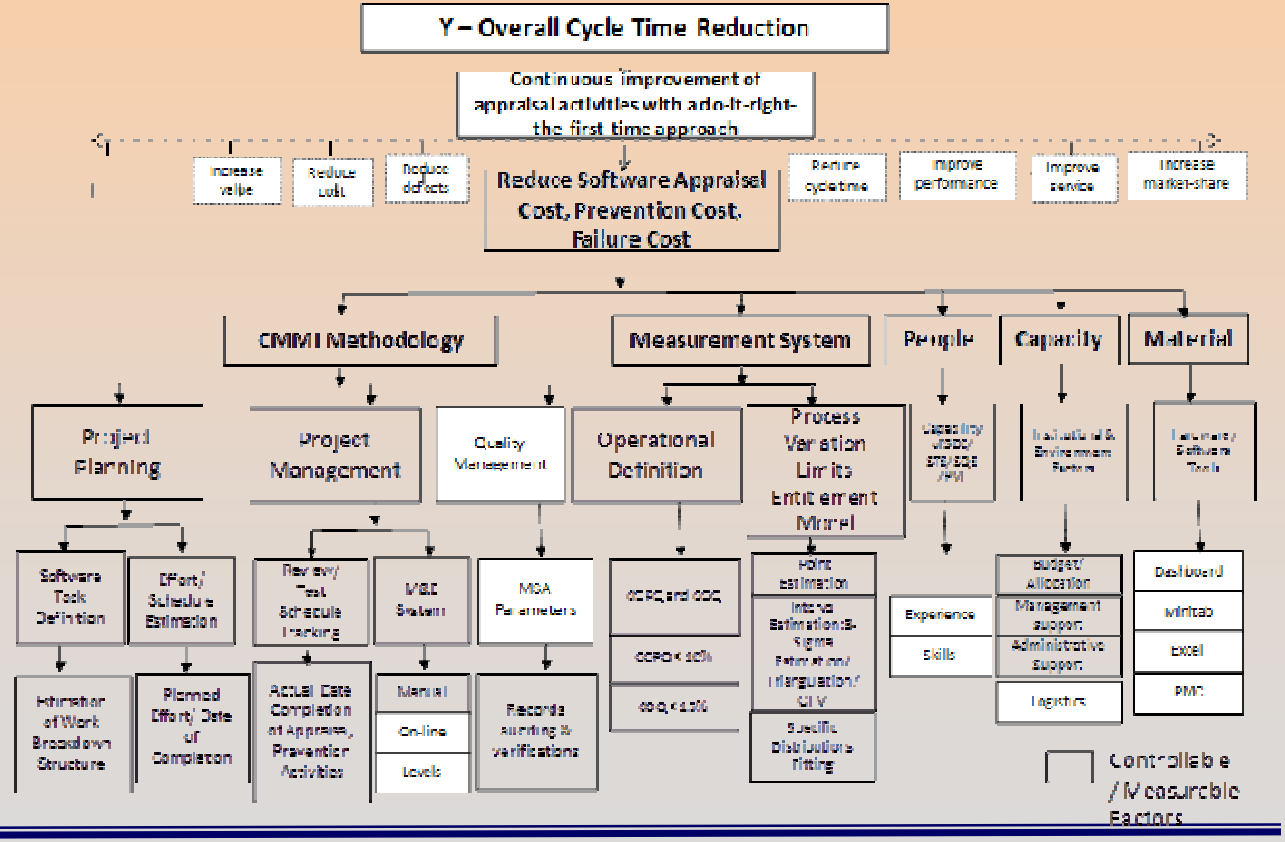
- Do we need this measure?
- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- **5 Other examples: CMMI-DEV Specific**
- SITARA TQI for TQA^{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A

Five Other Examples: CMMI-DEV Specific

(1 of 3)

Example 1: Cost of Quality

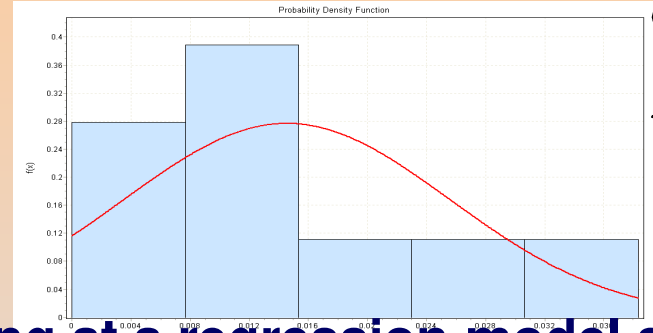
Step 1. Start, by representing business objectives in a Y-to-X Tree



Five Other Examples: CMMI-DEV Specific

(2 of 3)

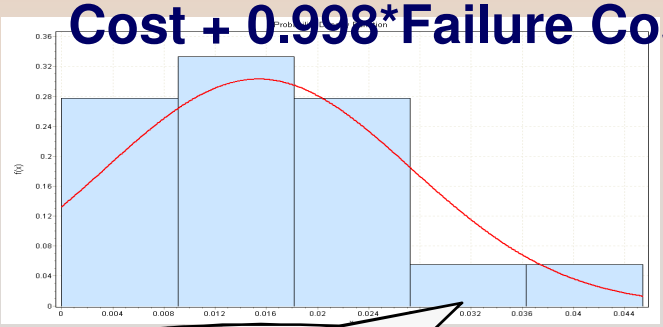
Step 2. Use the underlying probability density functions for the three predictors of …



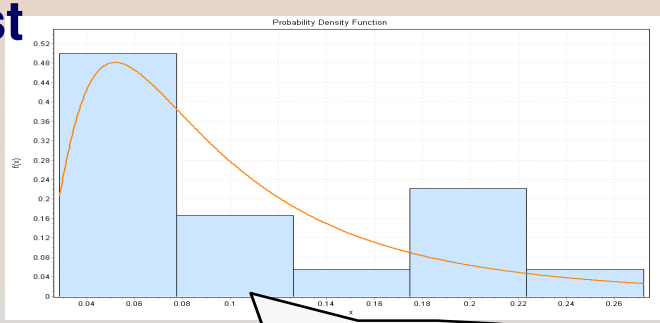
PDF: Appraisal Cost

… before arriving at a regression model such as:

COQ = 0.147 + 1.04*Appraisal Cost + 0.679*Prevention Cost + 0.998*Failure Cost



PDF: Prevention Cost

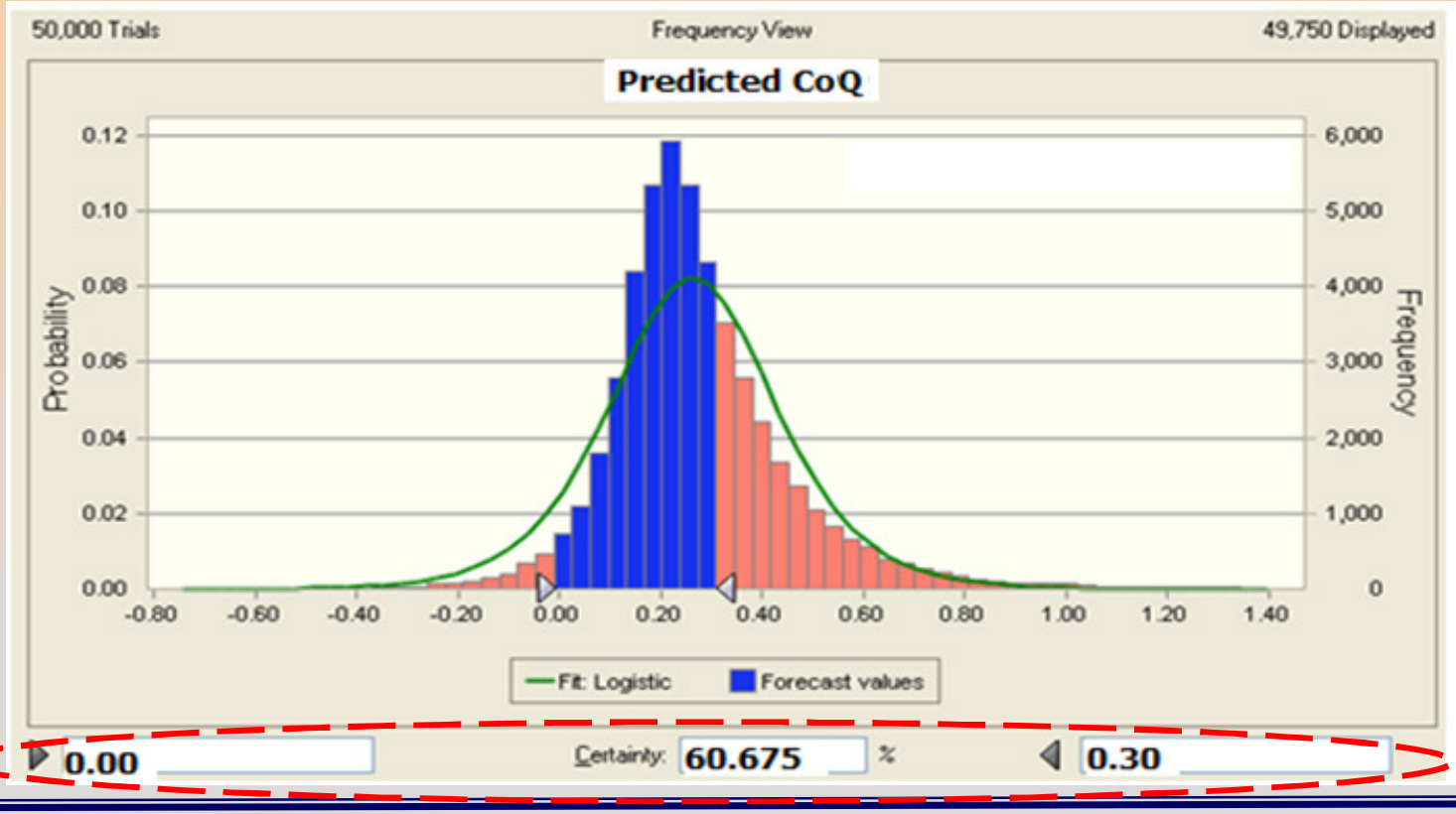


PDF: Failure Cost

Five Other Examples: CMMI-DEV Specific

(3 of 3)

Step 3. Perform what-if analysis and determine certainty levels of predicting CoQ to be between 0% and 30%…

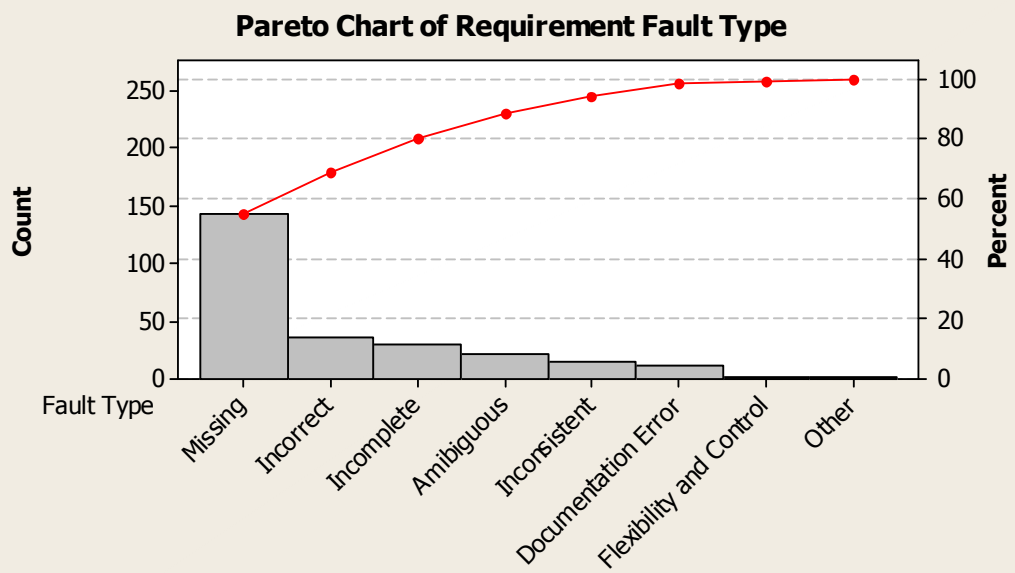


Five Other Examples: CMMI-DEV Specific

(1 of 2)

Example 2: Requirements Fault type Pareto Chart

Step 1. When more than one cause exists, conduct a one way ANOVA to test-for-differences, among equally likely causes



Count	142	36	30	22	14	12	2	2
Percent	54.6	13.8	11.5	8.5	5.4	4.6	0.8	0.8
Cum %	54.6	68.5	80.0	88.5	93.8	98.5	99.2	100.0

Five Other Examples: CMMI-DEV Specific

(2 of 2)

Step 2. Use Hypothesis Testing, to conclude existence of difference or otherwise of mean fault density for the different types

- Ho p >0.05 : Mean fault density by category is "="
- Ha p <0.05 : At least 1 Mean fault density by category is "≠"

One-way ANOVA: Missing Reqt., Incomplete Reqt., Incorrect Reqt., Documentation Error, Ambiguous Reqt. Inconsistent Reqt.

Source	DF	SS	MS	F	P
Factor	5	57.31	11.46	2.94	0.028
Error	31	120.76	3.90		
Total	36	178.08			

S = 1.974 R-Sq = 32.18% R-Sq(adj) = 21.25%

Level	N	Mean	StDev
Missing Requirement	7	3.420	3.763
Incomplete requirement	10	0.063	0.117
Incorrect requirement	6	0.050	0.067
Documentation Error	4	1.532	2.316
Ambiguous Requirements	5	0.877	0.826
Inconsistent requirement	5	1.645	2.051

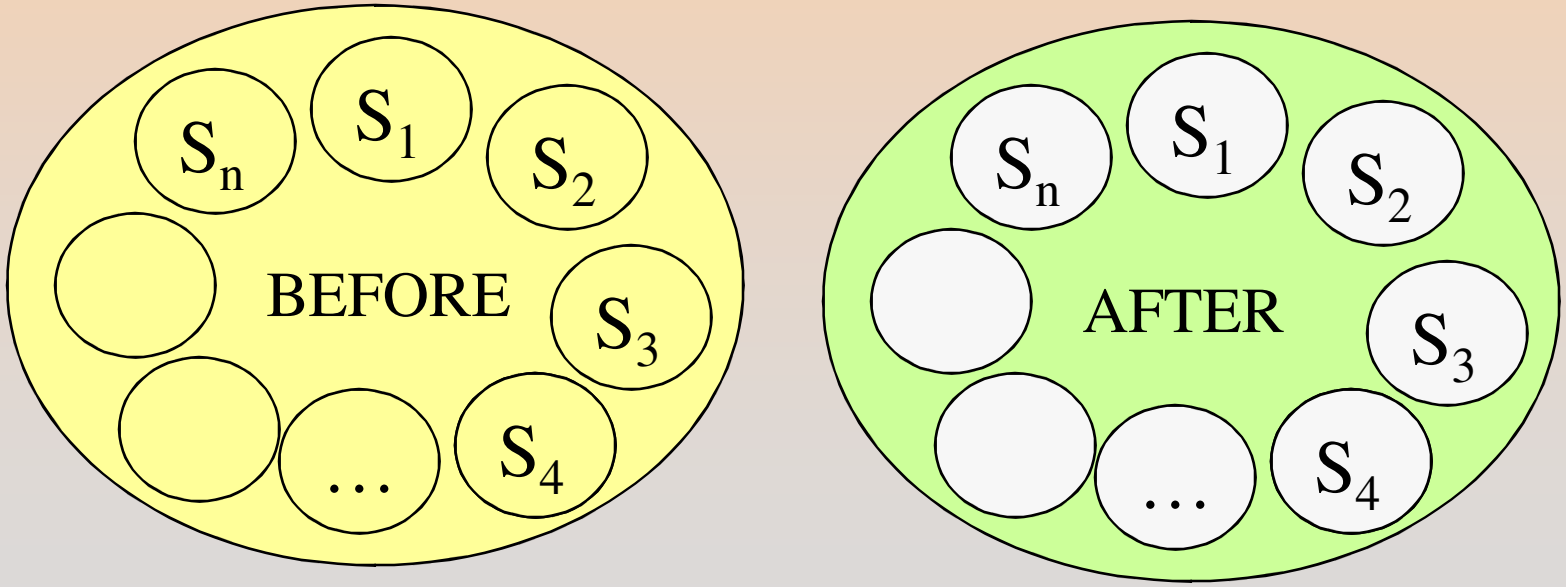
```
                            Individual 95% CIs For Mean Based on
                            Pooled StDev
Level                       --------+---------+---------+---------+-
Missing Requirement                            (-------*-------)
Incomplete requirement        (-----*------)
Incorrect  requirement      (-------*-------)
Documentation Error               (---------*---------)
Ambiguous Requirements          (--------*--------)
Inconsistent requirement          (--------*--------)
                            --------+---------+---------+---------+-
                                   0.0       2.0       4.0       6.0

Pooled StDev = 1.974
```

Five Other Examples: CMMI-DEV Specific

Example 3: Use of two sample T-Tests

Step 1. Collect adequate number of samples of data concerning the process parameter "before" and "after" say, a process improvement or innovation was introduced from a *homogenous* group of projects



Five Other Examples: CMMI-DEV Specific

Example 3: Use of two sample T-Tests *(contd. …)*

Step 2. Compute the average, standard deviation of the samples and setup the hypothesis for "Job Management Cycle Time"

Two-sample T

Ho: Job Management mean BEFORE improvement <= Job Management mean AFTER improvement

Ha: Job Management mean BEFORE improvement > Job Management mean AFTER improvement

```
Two-Sample T-Test and CI: Job Management (Before), Job Management (After)

Two-Sample T for Job Management *(Before) vs Job Management (After)

                                N      Mean     StDev   SE Mean
Job Management (Before)        10      3732       512       162
Job Management (After)         10      2241       752       238
```

Five Other Examples: CMMI-DEV Specific

(3 of 4)

Example 3: Use of two sample T-Tests *(contd. …)*

Step 3. Use the P-Value to determine statistical significance at say 95% confidence

Step 4. Draw conclusions about the effect of the process improvement or the innovation

```
Difference = mu (Job Management (Before)) - mu (Job Management (After))
Estimate for difference: 1491
95% lower bound for difference: 986
T-Test of difference = 0 (vs >): T-Value = 5.18 P-Value = 0.000 DF = 9
```

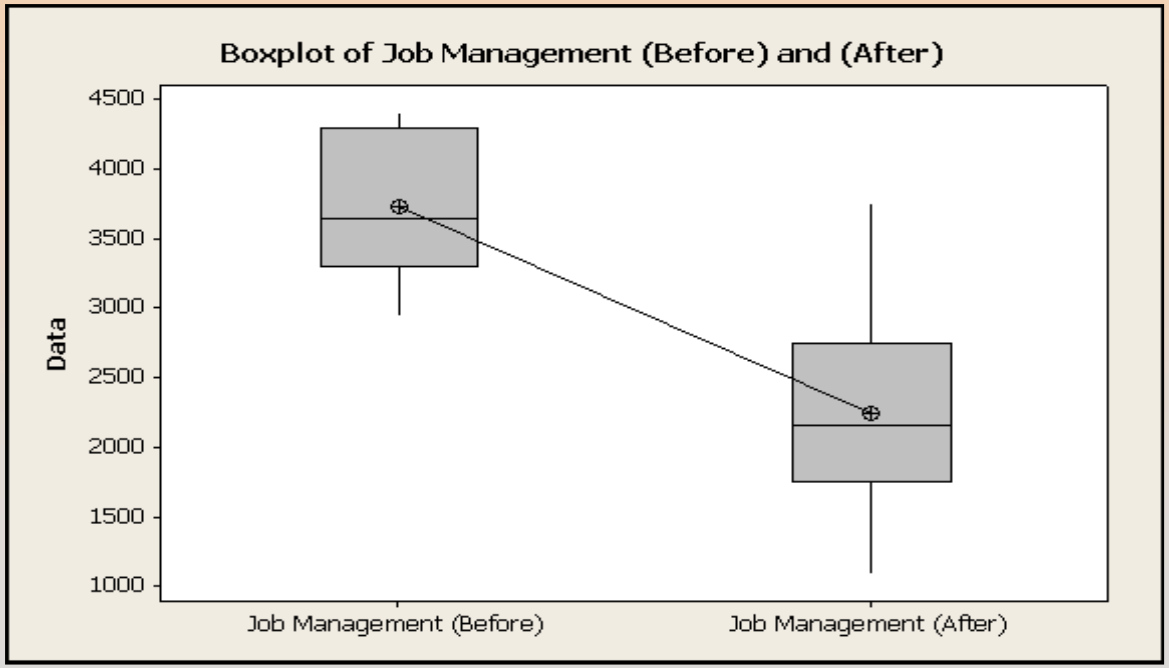
P value = 0.000	Reject Null Hypothesis
Conclusion	• Job Management cycle time mean after improvement is significantly less than before improvement.

Five Other Examples: CMMI-DEV Specific

(4 of 4)

Example 3: Use of two sample T-Tests *(contd. …)*

Step 5. Validate the Hypothesis using say, Box-plots; mean was brought down by 1,491.74 ms after performance improvement

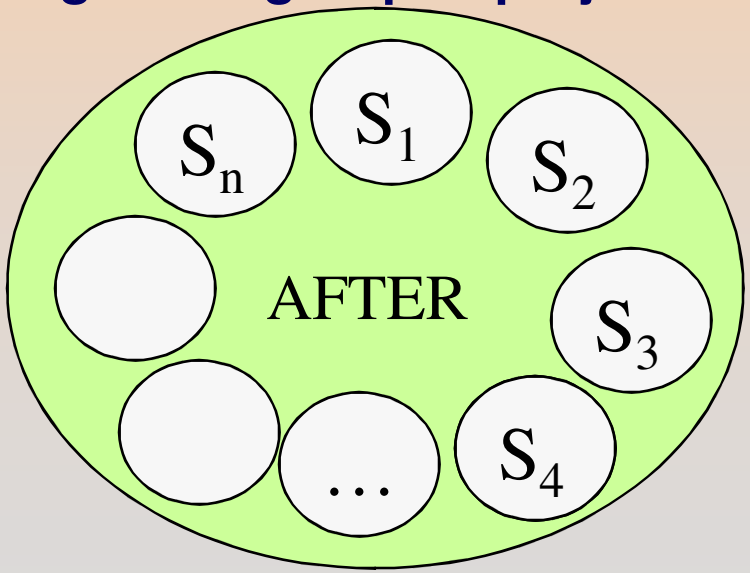


Five Other Examples: CMMI-DEV Specific

(1 of 2)

Example 4: Use Z-statistic (Classical Method) when large samples are available

Step 1. Collect adequate number of samples of data concerning the process parameter say, after a process improvement or innovation was introduced from a homogenous group of projects



Sample Mean = $\bar{x}$

Sample SD = s

Sample Size = n

Five Other Examples: CMMI-DEV Specific

(2 of 2)

Example 4: Use Z-statistic (Classical Method) when large samples are available *(contd. …)*

Step 2. Compute the Z-Statistic and set up the hypothesis for a target or goal value

$$Z = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$$

Organizational Goal

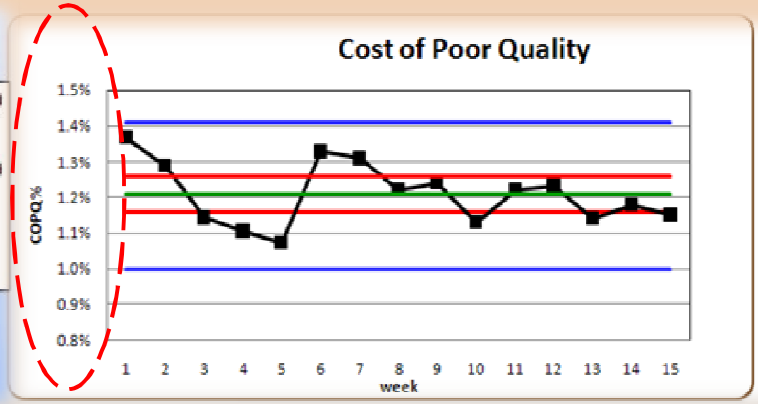
Step 3. If the Z value is <-1.96 or >1.96 (rejection region) for a level of significance = 0.05 or 95% confidence, then the null hypothesis would be rejected

Step 4. Draw conclusions about the effect of the process improvement or the innovation

Five Other Examples: CMMI-DEV Specific

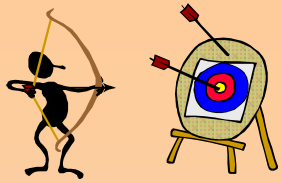
(1 of 1)

Example 5: Use control charts and 95% confidence interval for the "mean" to examine how the process has "settled" over time



Key point:

- every stable and capable process can be made more stable and more capable



- Do we need this measure?
- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- **SITARA TQI for TQA^{sm}**
  - Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A

SITARA Ten Question Indicator for Total Quality Assurance^{sm}

High-maturity Probe for Self-Assessment

(1 of 5)

1. What was a new idea/improvement we tried this month on our project?
2. Can we offer instances of how this improvement has led to demonstrate quantitative benefits on the process/product measures used within the organization?
3. What new measures do I need, that are not already available, for me to conduct my project better, faster, cheaper?
4. Can we demonstrate using 'test of hypothesis' or other statistical techniques how the project's measures are in alignment with the organization-level process capability baseline with respect to central tendency and variance?
5. What are the types of innovations we can adopt and institutionalize in our projects to demonstrate cycle time reduction or defect reduction?
6. What effects do decisions have, on effective deployment of causal analysis practices?
7. What has been the effect of risk management on cycle time reduction, defect prevention and customer satisfaction?
8. What steps do we have in place for the project to identify and take corrective actions even before the customer notices it?
9. What are some of the project-specific practices with which we can 'delight our customer'? Are they part of the standard operating practices?
10. What aspects of the project have been quantified and stabilized with which the project can demonstrate prediction of process capability?

SITARA TQI for TQA^{sm} – Sample Answers

(2 of 5)

"Changing requirements often impact most software projects. These changes can be due to internal or, external reasons. **Requirements Volatility**, for instance establishes a meaningful basis to establish stability of project requirements."

" Since effort and therefore, time is spent in the lifecycle phases, **phase-wise schedule deviation and phase-wise effort deviation** are useful metrics to have."

SITARA TQI for TQA^{sm} – Sample Answers

(3 of 5)

“ Two key measures that are useful in understanding process overheads (due to phase-wise rework effort and, phase-wise appraisal, prevention and failure efforts) are **cost of poor quality and cost of quality**.”

“ Since the deliverable that truly matters in software is a fully functional debugged-code, it is important to establish **Code Review Effectiveness** to understand the effectiveness of a review or a walkthrough.”

SITARA TQI for TQA^{sm} – Sample Answers

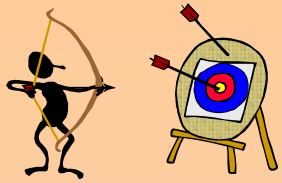
(4 of 5)

“ Every process improvement program must help in building a learning orientation for the organization. So, **phase-wise defect removal effectiveness** as a lag indicator that helps to address process limitations and improve defect removal in the different lifecycle phases ‘the next time around’ is a useful measure.”

SITARA TQI for TQA^{sm} – Sample Answers

(5 of 5)

"By the same token, a lead indicator for just how much time did the project consume in a phase in comparison to the total available project effort? – is yet another important question that is likely to be useful to understand. **Effort-to-time ratio** is a useful measure to have as a leading indicator of the effort already expended to the total project time available."



- Do we need this measure?
- How do we decide the most appropriate measure?
- Examples of –
  - Regression model for Interrelating Data
  - Summarizing Data
  - Comparing Data
- Other examples: CMMI-DEV Specific
- SITARA TQI for TQAsm
  - Ten Question Indicator for Total Quality Assurance with sample answers
- **Conclusion, Q & A**

Conclusion

Application of Quantitative and Statistical Methods in Software Engineering is primarily for 3 reasons –

- Summarizing Data,
- Comparing Data and,
- Interrelating Data

Based on the business objectives, quantitative and statistical methods in software engineering must consider –

- both, process factors and people factors,
- that contribute to minimizing process variation, and improving organizational learning
  - E.g., Conduct Causal Analysis on "successful outcomes" to establish the "success factors"

Conclusion

We understood how to decide the most appropriate measure for statistical and/or quantitative analysis

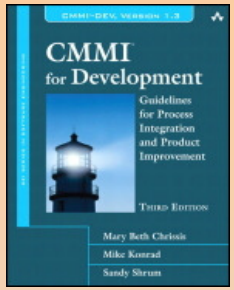
We reviewed some examples from high-process maturity implementations based on answers to the SITARA TQA for TQI^{sm} high-maturity probe for –

- Regression model for Interrelating Data
- Summarizing Data
- Comparing Data

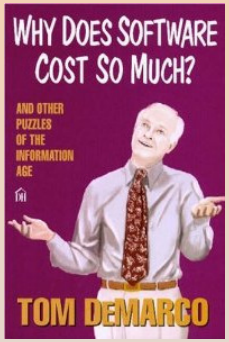
Other examples included –

- Simulations using "What-if Analysis"
- Hypothesis Testing :: One-way ANOVA, 2 sample T-Test and the Z-Statistic

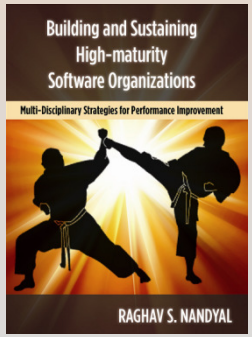
References



[Chrissis 2011] Chrissis, Mary Beth et al, "CMMI for Development®: Guidelines for Process Integration and Product Improvement, Third Edition", Addison-Wesley Professional. 2011.



[DeMarco 1995] "Why Does Software Cost so Much?"., T. DeMarco, Dorset House Publishing. 1995.



[Nandyal 2012] Nandyal, Raghav, "Building and Sustaining High-maturity Software Organizations", Tata McGraw-Hill Education. 2012.

Thank You!

DISCUSSIONS, Q&A

Raghav S. Nandyal
Chief Executive Officer

SITARA Technologies Pvt. Ltd.
#54, Sri Hari Krupa
6th Main Road
Malleswaram
Bangalore KA 560 003
Telephone: +(91-80) 2334-3222
Mobile: + 984-523-3222
Email: raghav_nandyal@SITARATECH.com
URL: http://www.SITARATECH.com