Examples of Application of Quantitative and Statistical Methods in Software Engineering

Raghav S Nandyal
Chief Executive Officer
raghav_nandyal@SITARATECH.com
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 ...

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
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
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.
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
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\textsuperscript{sm} – Ten Question Indicator for Total Quality Assurance with sample answers
- Conclusion, Q & A
Establish the Context

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<sup>sm</sup>), 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

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<sup>sm</sup> – Ten Question Indicator for Total Quality Assurance with sample answers

Conclusion, Q & A
Do we need this measure?

You need data to see what is inside the process; a lot of data!
Do we need this measure?

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?

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<sub>sm</sub> – Ten Question Indicator for Total Quality Assurance with sample answers

Conclusion, Q & A
How do we decide the most appropriate measure?

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 Technologies Pvt. Ltd.
For Use and Distribution by the SEI in the proceedings of the 24th SEPG Conference 2012 – Albuquerque
How do we decide the most appropriate measure?

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

- **Deterministic Model:** $F = \text{Mass} \times \text{Acceleration}$
  - Holds true, if the physics behind this equation has not changed

- **Probabilistic Model:** $\text{COQ} = 0.147 + 1.04\times\text{Appraisal Cost} + 0.679\times\text{Prevention Cost} + 0.998\times\text{Failure Cost} + e$
  - 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?
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?

**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<sup>sm</sup> – Ten Question Indicator for Total Quality Assurance with sample answers

Conclusion, Q & A
Example of a Regression Model

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

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.147</td>
<td>0.2227</td>
<td>0.66</td>
<td>0.538</td>
</tr>
<tr>
<td>Reqt. COPQ</td>
<td>1.03542</td>
<td>0.04228</td>
<td>24.49</td>
<td>0.000</td>
</tr>
<tr>
<td>Design COPQ</td>
<td>0.67922</td>
<td>0.03940</td>
<td>17.24</td>
<td>0.000</td>
</tr>
<tr>
<td>CUT COPQ</td>
<td>0.998249</td>
<td>0.009767</td>
<td>102.20</td>
<td>0.000</td>
</tr>
<tr>
<td>Test COPQ</td>
<td>1.00440</td>
<td>0.00756</td>
<td>132.90</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>94.371</td>
<td>23.593</td>
<td>511176.82</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>94.371</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reqt. COPQ</td>
<td>1</td>
<td>43.804</td>
</tr>
<tr>
<td>Design COPQ</td>
<td>1</td>
<td>49.715</td>
</tr>
<tr>
<td>CUT COPQ</td>
<td>1</td>
<td>0.036</td>
</tr>
<tr>
<td>Test COPQ</td>
<td>1</td>
<td>0.815</td>
</tr>
</tbody>
</table>

Interrelating Data
# Example of Summarizing Data

<table>
<thead>
<tr>
<th>Metric</th>
<th>Unit of measure</th>
<th>Q1 YYYY performance</th>
<th>Q2 YYYY performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>COPQ</td>
<td>%</td>
<td>10.66*</td>
<td>0.58</td>
</tr>
<tr>
<td>COQ Appraisal</td>
<td>%</td>
<td>16.88</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*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

Cost of Quality Before/After using Phase-Preview Checklist

**BEFORE**
- $\bar{x}=9.17$
- $UCL=16.08$
- $LCL=2.25$

**AFTER**
- $\bar{x}=5.17$
- $UCL=9.42$
- $LCL=0.91$
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<sup>sm</sup> – Ten Question Indicator for Total Quality Assurance with sample answers

Conclusion, Q & A
Five Other Examples: CMMI-DEV Specific

Example 1: Cost of Quality

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

Y - Overall Cycle Time Reduction

- Continuous improvement of appraisal activities with a don’t-right-the-first-time approach
- Reduce Software Appraisal Cost, Prevention Cost, Failure Cost

Y-to-X Tree Diagram:

- CMMI Methodology
- Measurement System
- Process Version Limits Enforcement Matrix
- People
- Capacity
- Material

Project Planning
- Software Risk Definition
- Status/Status of Completion

Project Management
- Quality Management
- Project Management

Operational Definition
- Process Version Limit Enforcement
- People
- Capacity
- Material

© SITARA Technologies Pvt. Ltd.
For Use and Distribution by the SEI in the proceedings of the 24th SEPG Conference 2012 – Albuquerque
Five Other Examples: CMMI-DEV Specific

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

... before arriving at a regression model such as:

\[ COQ = 0.147 + 1.04 \times \text{Appraisal Cost} + 0.679 \times \text{Prevention Cost} + 0.998 \times \text{Failure Cost} \]
Five Other Examples: CMMI-DEV Specific

Step 3. Perform what-if analysis and determine certainty levels of predicting CoQ to be between 0% and 30%...
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.
Five Other Examples: CMMI-DEV Specific

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 “≠”

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>5</td>
<td>57.31</td>
<td>11.46</td>
<td>2.94</td>
<td>0.028</td>
</tr>
<tr>
<td>Error</td>
<td>31</td>
<td>120.76</td>
<td>3.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>178.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Requirement</td>
<td>7</td>
<td>3.420</td>
<td>3.763</td>
</tr>
<tr>
<td>Incomplete requirement</td>
<td>10</td>
<td>0.063</td>
<td>0.117</td>
</tr>
<tr>
<td>Incorrect requirement</td>
<td>6</td>
<td>0.050</td>
<td>0.067</td>
</tr>
<tr>
<td>Documentation Error</td>
<td>4</td>
<td>1.532</td>
<td>2.316</td>
</tr>
<tr>
<td>Ambiguous Requirements</td>
<td>5</td>
<td>0.877</td>
<td>0.826</td>
</tr>
<tr>
<td>Inconsistent requirement</td>
<td>5</td>
<td>1.645</td>
<td>2.051</td>
</tr>
</tbody>
</table>

Individual 95% CIs For Mean Based on Pooled StDev

| Level                                | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
|--------------------------------------|---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Missing Requirement                  | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Incomplete requirement               | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Incorrect requirement                | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Documentation Error                   | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Ambiguous Requirements               | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
| Inconsistent requirement              | --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|

Pooled StDev = 1.974
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.
### 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”

<table>
<thead>
<tr>
<th>Two-sample T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ho:</strong> Job Management mean BEFORE improvement &lt;= Job Management mean AFTER improvement</td>
</tr>
<tr>
<td><strong>Ha:</strong> Job Management mean BEFORE improvement &gt; Job Management mean AFTER improvement</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Management (Before)</td>
<td>10</td>
<td>3732</td>
<td>512</td>
<td>162</td>
</tr>
<tr>
<td>Job Management (After)</td>
<td>10</td>
<td>2241</td>
<td>752</td>
<td>238</td>
</tr>
</tbody>
</table>
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

\[
\text{Difference} = \mu \text{(Job Management (Before))} - \mu \text{(Job Management (After))}
\]

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

<table>
<thead>
<tr>
<th>P value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>Reject Null Hypothesis</td>
</tr>
</tbody>
</table>

* Job Management cycle time mean after improvement is significantly less than before improvement.
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.

![Boxplot of Job Management (Before) and (After)]
Five Other Examples: CMMI-DEV Specific

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 = $\frac{\overline{X}}{S_n}$
Sample SD = $S_n$
Sample Size = $n$
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

\[
Z = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}
\]

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

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\textsuperscript{sm}**
- Ten Question Indicator for Total Quality Assurance with sample answers

Conclusion, Q & A
SITARA Ten Question Indicator for Total Quality Assurance

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?
“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.”
"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."
“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.”
“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\textsuperscript{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


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