Empirical Study of Software Engineering Results

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

TSP Initiative, SEAP/SDD
Sept. 18, 2013
If you are a data lover …

… then this interactive is for you.
If you are a data-lover, don’t be shy … 

… because you are in great company.
“In God we trust—all others bring data.”

W. Edwards Deming
“When it comes to the really important decisions, data trumps intuition every time.”

Jeff Bezos
That's 12 quadrillion bytes… or $10^{15}$ bytes.

12 petabytes for data storage.

It's not about me.
Unless you're using evidence-based practices, I can't hear a word you're saying.
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
- Next Steps
Please ask questions!
And ... we will be asking you questions!
IT'S A QUESTION PARTY!

MY FAVORITE!
Looking for FRESH ideas ...
All comments are welcome!
Would you like to collaborate?
What is the single most important question that you would want to be addressed through the analysis of TSP data?
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
- Next Steps
What Is Data Provenance?

Provenance, from the French provenir, "to come from," refers to the chronology of the ownership, custody or location of a historical object.

Data provenance refers to a record trail that accounts for the origin of a piece of data (in a document, repository, or database) together with an explanation of how it got to the present place.

Data provenance assists scientists with tracking their data through all transformations, analyses, and interpretations.
Be one with the Data!

Master Jim at work ... prying the data from the tool.
The TSP Data Repository includes two archives of approximately 50,000 files.

- 20-25% of those are supporting documents for a launch or postmortem (e.g., presentations, analysis spreadsheets, lists, surveys, etc.).
- 75-80% of those are TSP team performance data files.

There are more than 50 different file formats.

- At this time, only 22 of the needed file import utilities have been developed.
- Only those that contain TSP cycle or postmortem data were processed.
- About 60% of Archive #1 fit the criteria and have been processed. Archive #2 has not been processed yet.

Tests were conducted to ensure that extracted data represented unique projects.
Analysis Based on Teams’ Composite Data

The data is composite team data; each record represents data that has been aggregated from individual team member data.
Statistical Analysis

Individual team member data

Team composite data

To Analysis
Outlier Analysis

Exploratory analysis yielded some outliers that were removed from some of the analyses.
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
- Next Steps
Project Duration (Calendar Days)

Frequency

Project Duration (Days)

Mean = 119.4
Median = 91.0
n = 113
Project Duration (Weeks)

- Mean = 16.9
- Median = 13.0
- n = 113

Frequency

Duration (Weeks)

- Mean = 16.9
- Median = 13.0
- n = 113
Team Size

Mean = 8.2
Median = 8.0
n = 111
What other types of “team and project characteristics” analyses would you find valuable?
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
  - Team and Project Characteristics
  - Product Size
    - Schedule Performance
    - Quality Indicators
- Next Steps
Actual Total Size

Mean = 24.0
Median = 3.8
n = 111

Frequency

Total Size (KLOC)
Actual Added and Modified Size

Frequency

Mean = 11.7
Median = 6.6
n = 112

Actual Added & Modified Code (KLOC)
Another View of the Size Data

From the data, calculate the

1. log of each value
2. mean (\(\bar{x}\)) of values from #1
3. standard deviation (s) of values from #1
4. exponent of values from #2 and #3.

Values of the relative size table:

<table>
<thead>
<tr>
<th></th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Small</td>
<td>(\bar{x} - 2s)</td>
</tr>
<tr>
<td>Small</td>
<td>(\bar{x} - s)</td>
</tr>
<tr>
<td>Medium</td>
<td>(\bar{x})</td>
</tr>
<tr>
<td>Large</td>
<td>(\bar{x} + s)</td>
</tr>
<tr>
<td>Very large</td>
<td>(\bar{x} + 2s)</td>
</tr>
</tbody>
</table>
Added & Modified Code – Relative Sizes

- **Relative Size**
  - VS
  - S
  - M
  - L
  - VL

- **Frequency**
  - 0
  - 5
  - 10
  - 15
  - 20
  - 25
  - 30
  - 35
  - 40
  - 45
  - 50
What would you like to see in terms of analyses associated with product size?
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
  - Team and Project Characteristics
  - Product Size
  - Schedule Performance
  - Quality Indicators
- Next Steps
What do you think is the average number of weekly task hours that teams are able to accomplish?
Mean Team Member Weekly Task Hours

Mean = 10.3
Median = 9.0
n = 111
Productivity

Frequency

Productivity (LOC/Hr)

Mean = 10.3
Median = 7.1
n = 112
Let’s Look At Some Scatter Plots

$r$ is a measure of the correlation between $x$ values and $y$ values.

Values of $r^2 \geq 0.5$ indicate a meaningful relationship.

Using linear regression:

$$y = mx + b$$

where:

- $m$ is the slope
- $b$ is the y-intercept

Values of $r^2 \geq 0.5$ indicate a meaningful relationship.
There Are A Few Rules …

As you can see, by late next month you’ll have over four dozen husbands. Better get a bulk rate on wedding cake.
Plan Task Hours Vs. Actual Task Hours

R² = 0.8038

n=113
Duration: Planned Weeks Vs. Actual Weeks

$R^2 = 0.7401$

$n = 113$
Actual Task Hours Vs. Added & Modified LOC

![Scatter plot showing the relationship between actual task hours and added & modified LOC. The correlation coefficient $R^2 = 0.3012$.](image)

$n = 113$
Team Size vs. Productivity

![Graph showing the relationship between team size and productivity. The graph includes a scatter plot with data points and a trend line. The R² value is 0.0297.](image)

- **Productivity (LOC/Hr)**
- **Team Size**
- **R² = 0.0297**
- **n = 111**
Actual Added & Modified LOC Per Staff Week

n = 109

Mean = 204.9
Median = 100.9
n = 109
Plan Vs. Actual Hours for Completed Parts

R² = 0.952

Actual Hours for Completed Parts vs. Plan Hours for Completed Parts

n = 113
Schedule Growth Beyond Baseline

Mean = 2.8
Median = 0.0
n = 113

Schedule Growth from Baseline (hours)

Frequency
Final Earned Value

Mean = 85.2
Median = 94.3
n = 110
Are there other types of “schedule performance” analyses that you would like to see?
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
  - Team and Project Characteristics
  - Product Size
  - Schedule Performance
  - Quality Indicators
- Next Steps
Total Defects Injected Per KLOC

Mean = 54.7
Median = 31.5
n = 79
Defect Density – DLD Review

Mean = 5.2
Median = 2.2
n = 106

Defects Per KLOC - DLD Review

Frequency

0 4 8 12 16 20 24 28

0 10 20 30 40
Defect Density – Code Review

Mean = 8.4
Median = 5.2
n = 106
Defect Density – Code Inspection

- Mean = 6.7
- Median = 3.3
- n = 106

Defects Per KLOC - Inspection

Frequency

0 10 20 30 40 50

0 20 40 60 80 100 120

Defects Per KLOC - Inspection
Defect Density – Unit Test

Mean = 5.0
Median = 3.8
n = 106

TSP Guidelines
< 5
Defect Density – Build and Integration Test

Mean = 1.7
Median = 0.7
n = 106

TSP Guidelines
< 0.5
Defect Density – System Test

- Mean = 1.5
- Median = 0.15
- n = 106

Defects Per KLOC - System Test
Defect Density - Summary

- Defects Per KLOC - DLD Review
- Defects Per KLOC - Code Review
- Defects Per KLOC - Inspection
- Defects Per KLOC - Unit Test
- Defects Per KLOC - Build and Integration Test
- Defects Per KLOC - System Test

Frequency
Defect Density – Median of Defects Per KLOC

- DLD Review: 2.2
- Code Review: 5.2
- Code Inspection: 3.3
- Unit Test: 3.8
- Build/Integration Test: 0.7
- System Test: 0.15

Defects Per KLOC (Median)
Actual UT Defect Density Vs. ST Defect Density

$n=111$

R² = 0.0031

System Test Defect Density vs. Unit Test Defect Density
The Yield Quality Measure

The yield of a phase is the percentage of defects removed in that phase.
Yield: Detailed-Level Design Review

Frequency

Yield - Detailed-Level Design Review

Mean = 31.6%
Median = 29.3%
n = 98
Yield: Detailed-Level Design Inspection

Mean = 49.3%
Median = 50.0%
n = 83
Yield: Code Review

Frequency

Yield - Code Review

Mean = 29.2%
Median = 30.1%
n = 109
Yield: Code Inspection

Mean = 30.3%
Median = 26.1%
n = 110

Yield - Code Inspection

Frequency

Yield: Code Inspection
Mean = 30.3%
Median = 26.1%
n = 110
Yield: Unit Test

- Mean = 49.7%
- Median = 46.2%
- n = 106
Summary: Median Phase Yields

- DLD Review: 29.3
- DLD Inspection: 50.0
- Code Review: 30.1
- Code Inspection: 26.1
- Unit Test: 46.2
Process Yield

Mean = 73.6%
Median = 73.7%
n = 77
Review of Some Definitions …

In the TSP:

\[
\text{Appraisal COQ} = \frac{\text{Review & Inspection Time}}{\text{Total Development Time}} \times 100
\]

\[
\text{Failure COQ} = \frac{\text{Test Time}}{\text{Total Development Time}} \times 100
\]
Appraisal Cost of Quality

Mean = 26.6%
Median = 27.0%
n = 86
Failure Cost of Quality

Mean = 22.0%
Median = 16.9%
n = 86
Appraisal: COQ vs. Defects Removed

\[ R^2 = 0.0347 \]

Defects Per KLOC Removed During Appraisal Phases

Appraisal Cost of Quality (Percent)

n=81
Time in Code Review Vs. Defects Removed

Defects Removed in Code Review vs. Actual Time in Code Review (Hours)

R² = 0.495

n=112
Time in Inspection Vs. Defects Removed

- Defects Removed During Inspection
- Actual Time in Inspection (Hours)

R² = 0.6272

n=112
Time in Unit Test Vs. Defects Removed

R² = 0.3505

n=112
Time in System Test Vs. Defects Removed

Actual Time in System Test (Hours)

Defects Found

R² = 0.1996

n=29
## Summary: Time in Phase Vs. Defects Removed

<table>
<thead>
<tr>
<th>Phase</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal</td>
<td></td>
</tr>
<tr>
<td>Code Review</td>
<td>0.50</td>
</tr>
<tr>
<td>Inspection</td>
<td>0.63</td>
</tr>
<tr>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Unit Test</td>
<td>0.35</td>
</tr>
<tr>
<td>System Test</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Predictive
Code Review to Code – Actual Time in Phase

- Mean = 0.3
- Median = 0.3
- n = 113

TSP GUIDELINES

> 0.5
Design to Code – Actual Time in Phase

Mean = 0.9
Median = 0.8
n = 113

TSP GUIDELINES
> 1.0

Design to Code - Actual Time
Design Review to Design – Actual Time in Phase

![Histogram showing the frequency of actual time from design review to design.]

- **Mean**: 0.28
- **Median**: 0.23
- **n**: 113

TSP Guidelines
- > 0.5
Req. Inspection to Req. – Actual Time in Phase

Mean = 0.88
Median = 0.46
n = 19

Frequency

TSP GUIDELINES

> 0.25
For “quality indicators,” what additional analyses would you like to see?
What is the single most important question that you would want addressed through the analysis of TSP data?
Topics

- Introduction
- Format of the “Interactive”
- Data Provenance
- The Data
  - Team and Project Characteristics
  - Product Size
  - Schedule Performance
  - Quality Indicators

Next Steps
WHAT'S NEXT?

- Review your feedback from today and adjust the analysis approach accordingly.
- Extract the data from Process Dashboard tool submitted files.
- Extract and analyze individual team member data.
- Continue with the data analysis. Publish the results.
Contact Information

Mark Kasunic  
Senior Member of Technical Staff  
TSP Initiative  
Telephone: +1 412-268-5863  
Email: info@sei.cmu.edu

U.S. Mail  
Software Engineering Institute  
Customer Relations  
4500 Fifth Avenue  
Pittsburgh, PA 15213-2612  
USA

Web  
www.sei.cmu.edu  
www.sei.cmu.edu/contact.cfm

Customer Relations  
Email: info@sei.cmu.edu  
Telephone: +1 412-268-5800  
SEI Phone: +1 412-268-5800  
SEI Fax: +1 412-268-6257