TOSHIBALeading Innovation >>>

SPIトワイライトフォーラム 2015年7月 ソフトウェア開発プロセスデータに関する 研究とSEI留学帰国便り

株式会社 東芝 インダストリアルICTソリューション社 IoTテクノロジーセンター プロセス・品質技術開発部 白井 保隆

2015年7月13日

私紹介

• 社内プロセス改善は、10年以上のキャリア

- コーポレートSEPG(全社レベルのSPI推進):10年くらい
- カンパニーSEPG(社会インフラ事業部のSPI推進):3年くらい

• JASPICにも参加していた

- SPI推進課題分科会、SPIBOK分科会
- 今度、JASPIC同窓会にお邪魔する予定です

留学

- 2013年9月~2015年3月
- カーネギーメロン大学ソフトウェア工学研究所 (CMU/SEI)
- 戻ってきてからはアジャイルで奮闘中!

目次

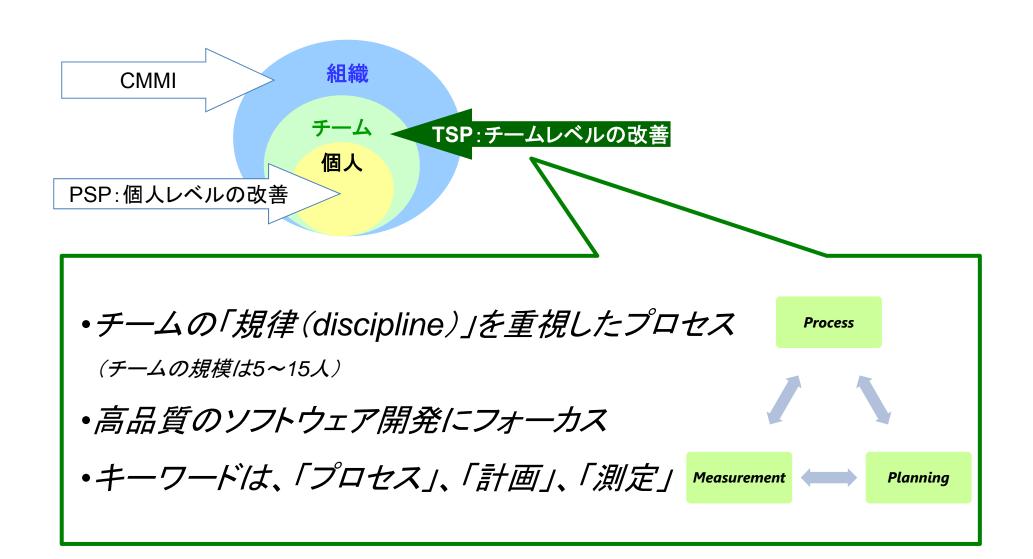
- 1. TSPとは
- 2. 研究のステップ
 - 2.1 研究の狙い
 - 2.2 TSPデータベースの構築
 - 2.3 データ品質評価
 - 2.4 データ分布調査
 - 2.5 品質モデルの構築
- 3. 研究のまとめとSEIの動向

目次

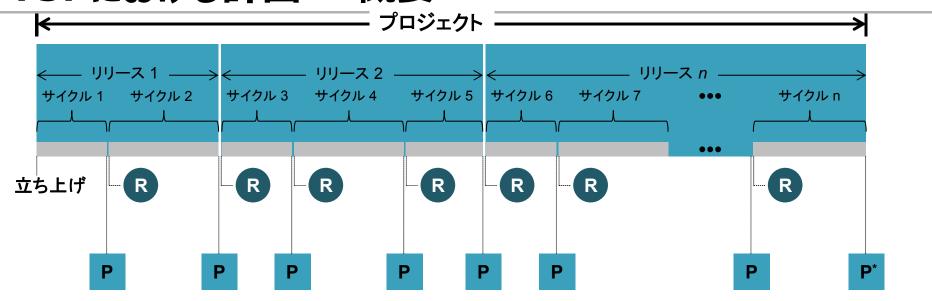
1. TSPとは

- 2. 研究のステップ
 - 2.1 研究の狙い
 - 2.2 TSPデータベースの構築
 - 2.3 データ品質評価
 - 2.4 データ分布調査
 - 2.5 品質モデルの構築
- 3. 研究のまとめとSEIの動向

TSP (Team Software Process) とは



TSPにおける計画 - 概要



最初の立ち上げ計画は、最後のリリースまでをカバーしている。

サイクルは計画を立案する期間。サイクルの最初には立ち上げ/再立ち上げを、最後にはポストモーテムを実施する。

- ・ 直近のサイクルは、活動を実施するチームによってより詳細な計画が立てられる。
- ・ 未来のサイクルは、立ち上げで計画されるが、直近のサイクルほど詳細なものではない。
- ・ 新しいサイクルは再立ち上げで始まる。

リリースやサイクルの期間はプロジェクトによって異なる。

立ち上げ重視!

P:ポストモーテム

R : 再立ち上げ

🏅 : プロジェクトポストモーテム

進捗の追跡とステータス報告

- 自己管理型のチームは、自らの計画を立案し、管理することに 責任を持つ。
- TSPプロジェクトでは、4つの基礎尺度を収集することで進捗を 追跡している。

スケジュール



規模



タスク時間



不具合



データ収集のためのツールサポート

- TSPソフトウェアツール(Software Process Dashboard etc)は、4つのコア尺度の収集をサポート
- データは、チームメンバの計画に組み込まれたタスクに沿って収集
- 個人の追跡データは、週次を基本としたチームプランに統合
- Software Process Dashboard は以下よりダウンロード可能



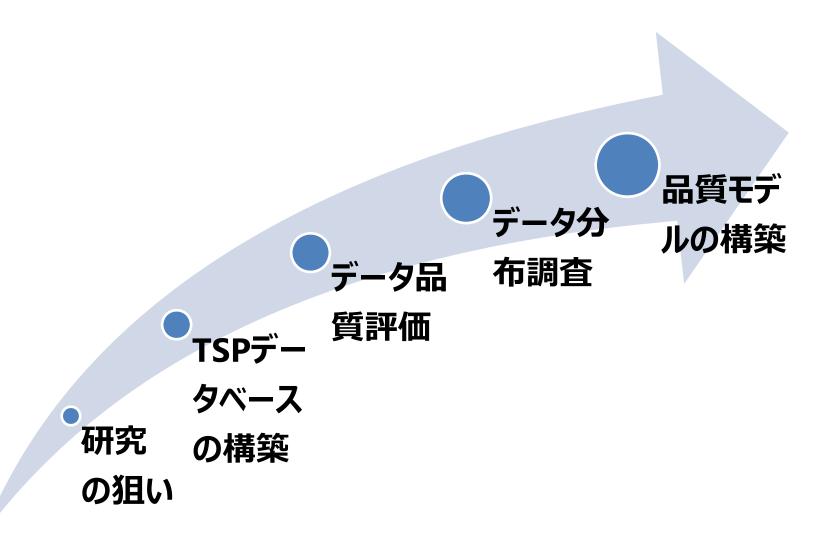
TSPで用いられるフォームの例

タスク	期間	中断時間	•	差	差分時間		∃付&タイ <i>Ь</i> √プ	スタ	コメント	
時間記録ログ										
									プウォッチ 目動カウン	
内容		欠陥型			除去 フェーズ	発見時間	見&修正 間	日付タンプ	& <i>タ</i> イムス プ	
欠陥記録ログ										
新規行		加と 削 正行 数	涂行	再利用 行数	合計行数		成果物規模 の単位	中	画 or 途 修正 or 績	
規模データ										

目次

- 1. TSPとは
- 2. 研究のステップ
 - 2.1 研究の狙い
 - 2.2 TSPデータベースの構築
 - 2.3 データ品質評価
 - 2.4 データ分布調査
 - 2.5 品質モデルの構築
- 3. 研究のまとめとSEIの動向

研究のステップ

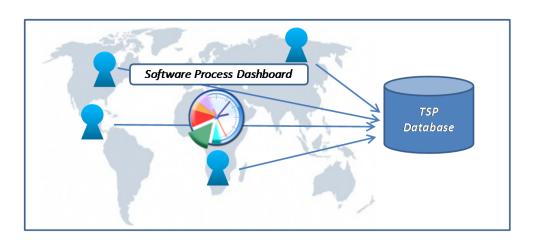


2.1 研究の狙い

- TSP適用プロジェクトのデータベースを構築、データ分析結果を公開することで、
 - 産業、研究でのデータ分析結果の活用促進を狙う
 - ベンチマーク、予測など
 - ソフトウェアエンジニアリング研究分野の発展を狙う

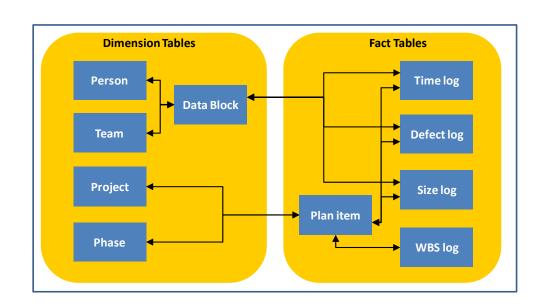
2.2 TSPデータベースの構築

- 通称:SEMPR
 - Software Engineering Measured and Performance Repository
- 109プロジェクトサイクル(2014年5月現在)
- Software Process Dashboardの情報を蓄積
- MySQL 5.6を使用



TSPデータベース

- •オープンソース「Team Process Data Warehouse」のスキーマを活用
 - http://www.processdash.com/tpdw
- •「Fact Tables」と「Dimension Tables」で構成
- ●「Fact Tables」と「Dimension Tables」により、様々な観点、様々な粒度での分析が可能



2.3 データ品質評価

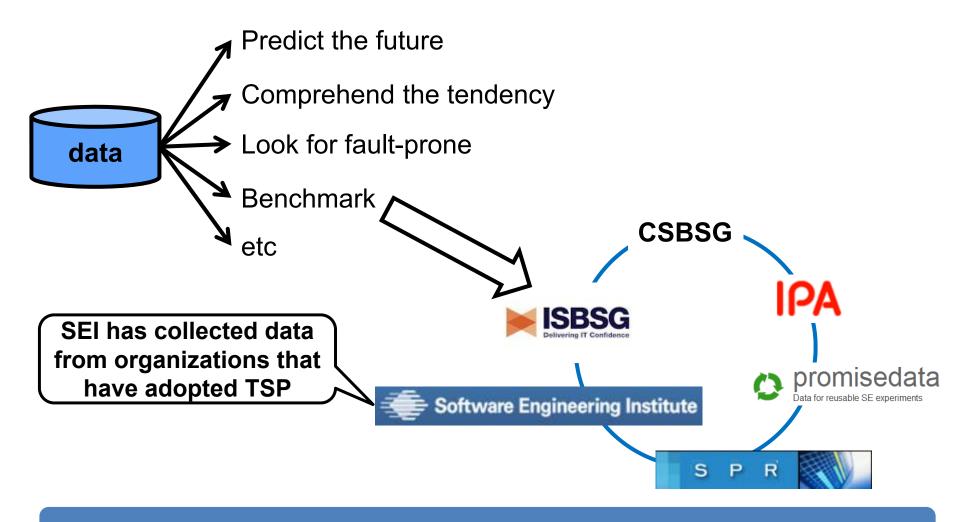
● 成果

- TSPプロジェクトデータ品質評価の枠組みを構築
- ICSSP2014(中国 南京)にて発表
- TSPプロジェクトデータの品質は
 - 時間と欠陥ログ: 品質はなかなか良い(自動でとれることが多い)
 - 規模とタスクログ: 品質はあまり(手動でとるところがいまひとつ)



Barry Boehm氏とICSSP2014にて

なぜデータ品質にフォーカスするのか



高いデータ品質の保証は、データ活用のためには必要不可欠

ISO/IEC 25012: データ品質モデルを活用

- Published in 2008
- Defined 15 quality attributes and 2 dimensions of inherent and system dependent

No	Attributes	Inherent	System Depende nt
<u>1</u>	<u>Accuracy</u>	X	
2	Completeness	X	
<u>3</u>	Consistency	X	
<u>4</u>	Credibility	X	
5	Currentness	X	
6	Accessibility	X	X
7	Compliance	X	X
8	Confidentiality	X	X

No	Attributes	Inherent	System Depende nt
9	Efficiency	X	X
10	Precision	X	X
11	Traceability	X	X
12	Understandabilit y	X	X
13	Availability		X
14	Portability		X
15	Recoverability		X

SEMPRにおけるデータ品質評価

●4つの必須データ品質属性にフォーカス

ID	Dimensions	Our verification item			
(1)	Amount of Data	- The number of records in the tables			
(2)	Correctness	- Illegal formats			
		- Missing data			
		- Wrong values			
(3)	Consistency	- Sequence consistency			
		- Effort duration consistency			
(4)	Credibility	- Direct tests			
		Overlapping time			
		Questionable work durations			
		- Indirect test			
		- Duration of Individual time record value			

(1) Amount of Data (2014年5月時点)

- •SEMPRのテーブル内のレコード数
- ・データ品質のエラー比率を算出する際に使用
- Dimension Tables
 - 109 projects
 - 34 teams
 - 309 people
- Fact Tables
 - See table

Fact Table	Amount of Data
Time Log	103,023
WBS Log	11,449
Defect Log	18,408
Size Log	7,464

(2) Correctness (正確性)

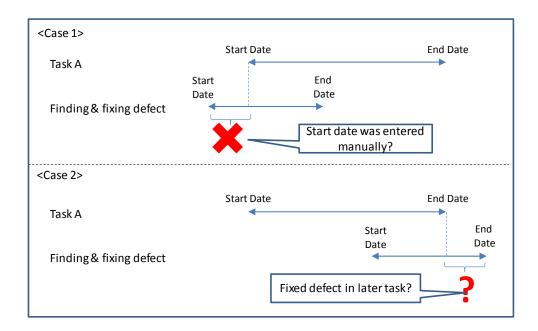
- 不正確なデータの数を調査
 - ルールを逸脱した形式
 - データ欠損
 - エラー値
- 自動収集可能なデータ(Time log、Defect log)の正確性は 高い
- Fix defect fieldとSize logは不正確なデータが多い(手動 入力のため)

Fact table	Amount of data	Illegal format	Lack of data	Wrong number	Incorrectness rate
Time log	103,023	1	2	0	0.003%
WBS log	11,449	0	0	0	0%
Defect log	18,408	0	442	277	3.9%
Fix defect field	828	0	442	0	53.4%
Size log	7,464	0	2,688	24	53%

(3) Consistency (一貫性)- Sequence consistency

- ●タスクと、該当タスクで出た不具合の発見と修正が、正しい順序で実行されているかを調査
 - ◆タスクと不具合の発見と修正の開始時刻、終了時刻にフォーカス

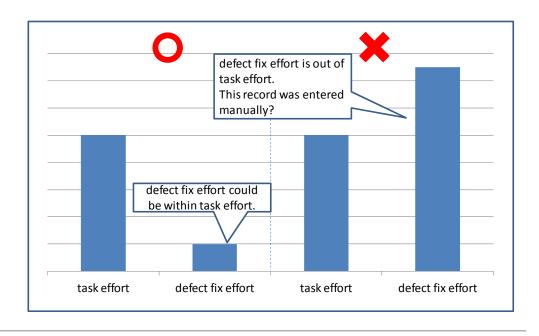
2.3% のタスクが一貫性がないことが判明



(3) Consistency (一貫性)- Effort Duration Consistency

●タスクの実行時間と、該当タスクで出た不具合の発見と修正の 実行時間の間に不整合がないかを調査

0.66% のタスクが、不具合の発見 と修正時間の方が多い事象が発 生していた



(4) Credibility (信頼性)

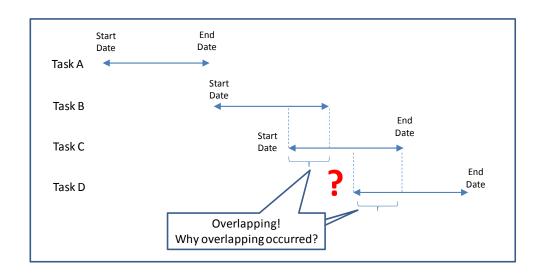
- ・「データが後で追加/修正されているか?」に着目して信頼性を評価
- ・直接的な評価
 - Time Overlapping
 - Questionable Work Times
- 間接的な評価
 - Duration of individual time record value
 - Distributional properties of the data

(4) Credibility (信頼性)- Time Overlapping

• タスク間の時間の「オーバーラップ」にフォーカス

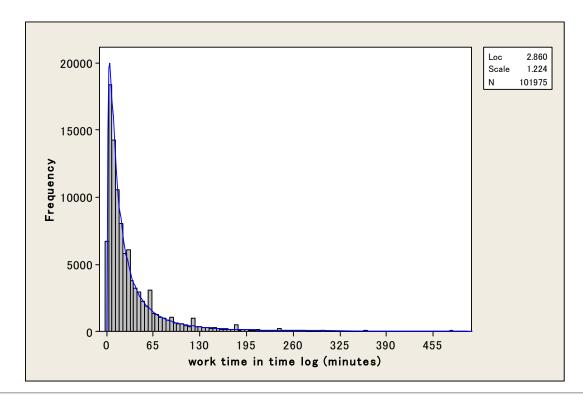
- ツールの特性上、時間のオーバーラップは、データの修正によって発生することに 着目

13.8%のタスクに時間のオーバーラップを発見



(4) Credibility (信頼性)-Duration of Individual Time Record Value

- ◆タスクの実行時間の分布に着目
 - ◆タスクの実行時間は、「対数正規分布」に従う、という仮説のもと、確率分 布から外れる実行時間を調査
- ●30分毎に「スパイク」が出来ていることを確認
 - ず一タが後で入力されている可能性が高い



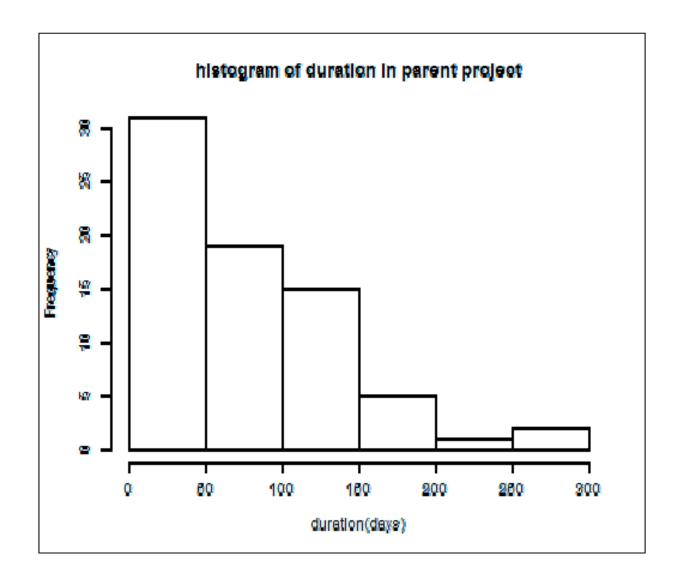
2.4 データ分布調査

● 成果

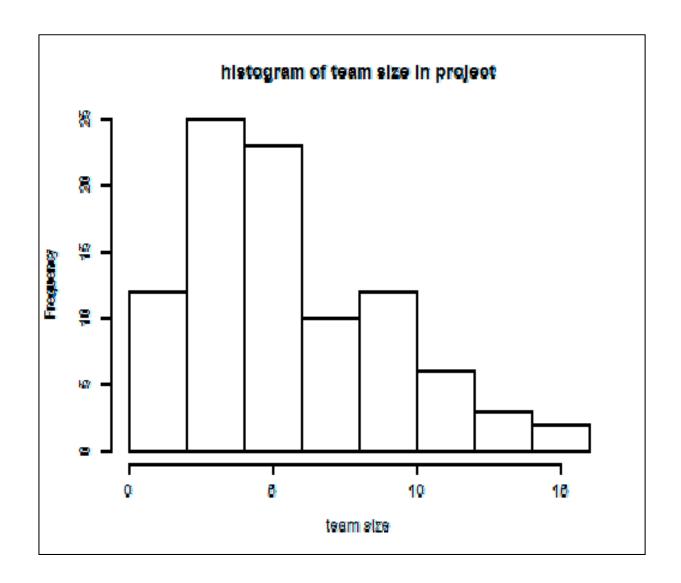
- 分布評価の一部は、SEI主催のTSPシンポジウム2014にて共同発表
- 今回、いくつかのデータ分析結果を紹介
 - プロジェクト(サイクル)全体を俯瞰した分析
 - 個別のプロジェクトに対する分析

プロジェクト全体を俯瞰した分析

SEMPRのプロジェクトサイクルの期間分布

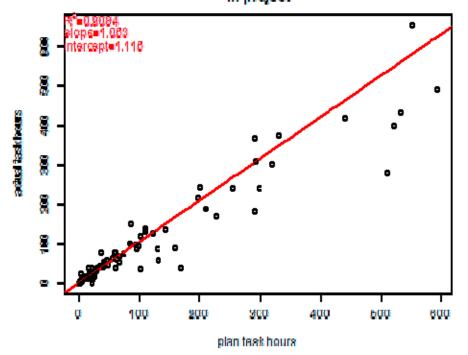


SEMPRのチームサイズの分布

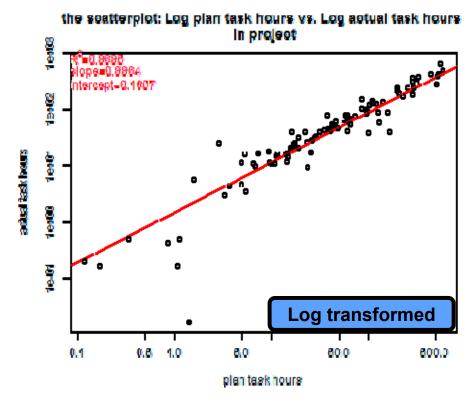


工数の予実(プロジェクトサイクル)

the scatterplot: Plan task hours vs. Actual task hours in project

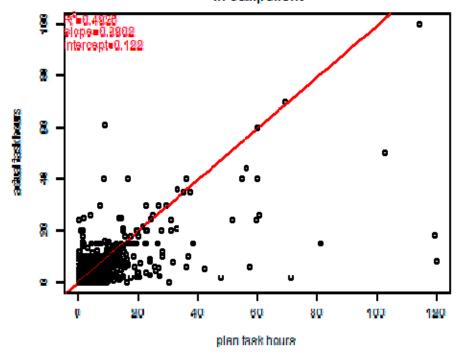


プロジェクトレベルの工数は予測可能性が非常に高い



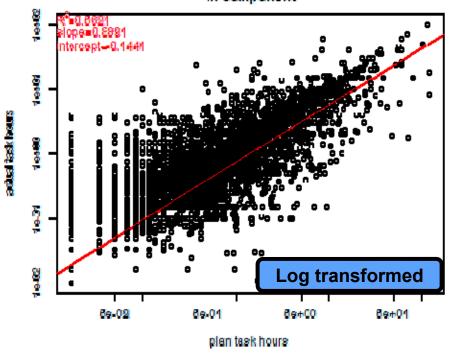
工数の予実(Work item)

the scatterplot: Plan task hours vs. Actual task hours in component

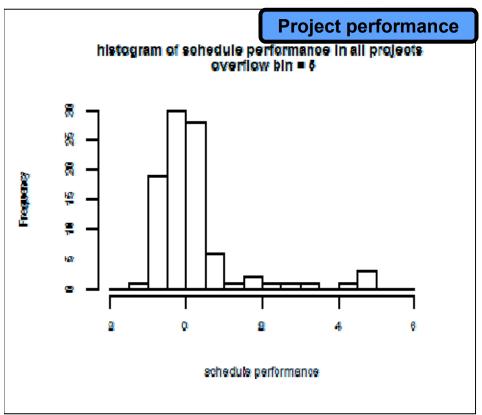


Work itemレベルの工数の対数は 予測可能である

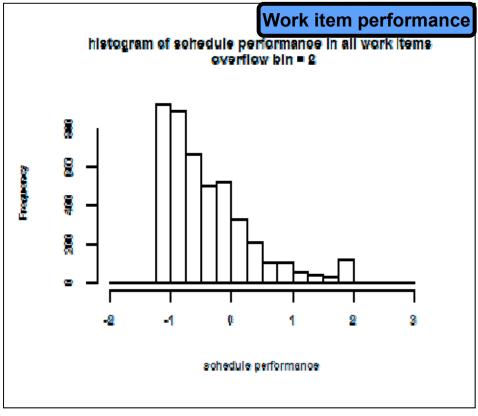
the scatterplot: Log plan task hours vs. Log actual task hours in component



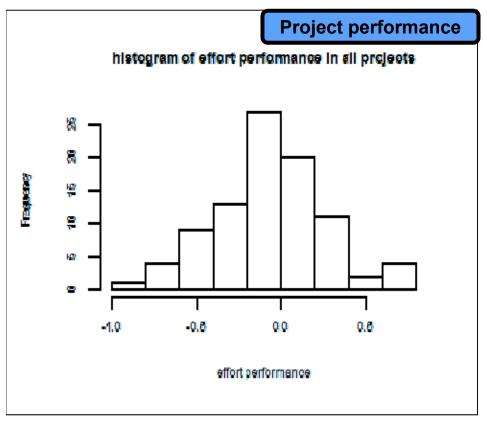
スケジュールパフォーマンス



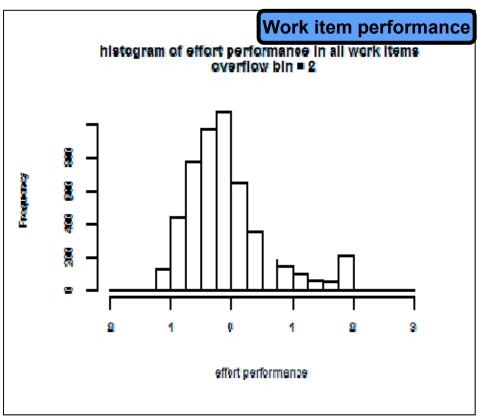
actual duration - plan duration plan duration



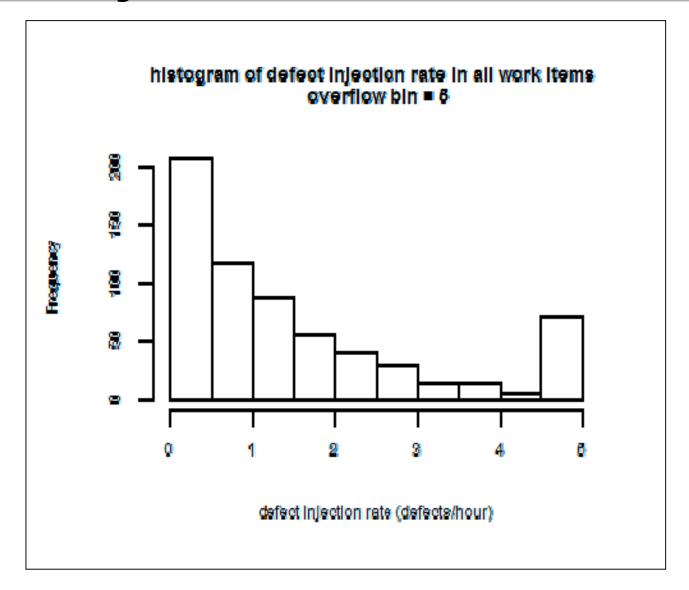
工数パフォーマンス



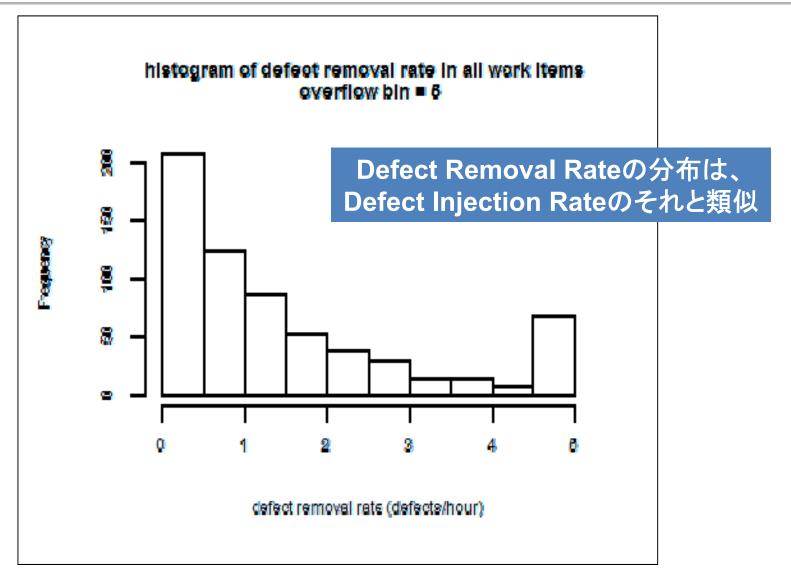
effort performance = actual effort - plan effort plan effort



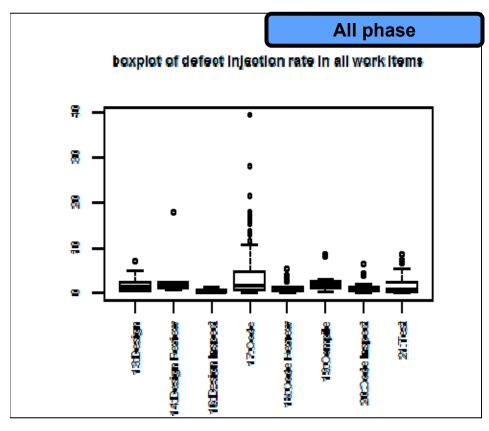
Defect Injection Rateの分布(Work item)



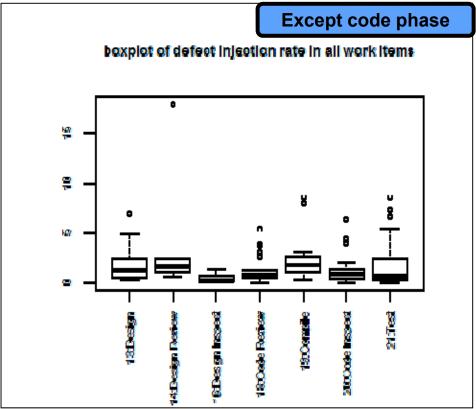
Defect Removal Rateの分布(Work item)



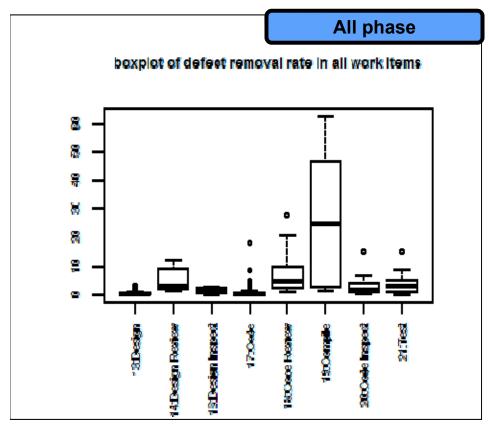
フェーズ毎のDefect Injection Rate



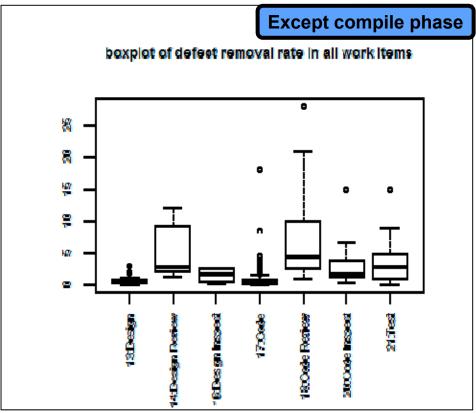
コードフェーズのDefect Injection Rateの分布は、広範囲かつ、もっとも 高い中央値を持つ



フェーズ毎のDefect Removal Rate



コンパイルとコードレビューでの Defect Removal Rateは、単体テストのそれよりも高い(中央値について)



個別のプロジェクトに対する分析

調査項目:記述統計

• 以下の統計について調査を実施

略称	名称	説明
N	サンプルサイズ	データの数
min	最小値	データの中で最も小さい値
max	最大値	データの中で最も大きい値
range	範囲	最大値から最小値を引いたもの
median	中央値	データを順に並べた際に、真ん中に来るもの
mean	平均值	すべてのデータの和をその総数で割ったもの
SE	標準誤差	データの平均値のバラつき (標準偏差を√Nで割ったもの)
CI	信頼区間	母集団全体の値がその区間に存在する確率が95%である区間のこと
Var	分散	データの平均値からのバラつきを表すもの
SD	標準偏差	データの平均値からの平均的なバラつきを表すもの
CV	変動係数	相対的なバラつきを表す (標準偏差を平均で割ったもの)
Q1, Q3	分位数	Q1: データを小さい順に並べたときの、小さい方から25%目の値 Q3: データを小さい順に並べたときの、小さい方から75%目の値

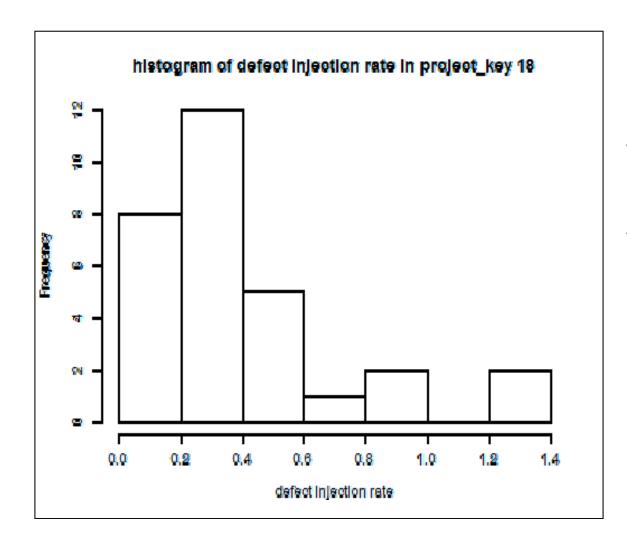
調査項目:分布のあてはまり

- データの分布が、確率分布にどれほど当てはまっているかを調査
 - 確率分布へ当てはまるかどうかが、そのデータが予測に使えるかどうかの鍵
 - 「対数正規分布」と「ワイバル分布」にフォーカス
- 分布のあてはまりの検証に利用できる手法
 - グラフィカルな手法: Q-Q plot, P-P plot
 - グラフィカルではない手法: アンダーソン・ダーリン検定
- 確率分布のパラメータ推定には、最尤推定法を利用

Defect injection rate

- Defect Injection Rate (以降、DIR)
 - = 不具合埋め込み数/工数(人時)

例:あるプロジェクトのDIRの記述統計



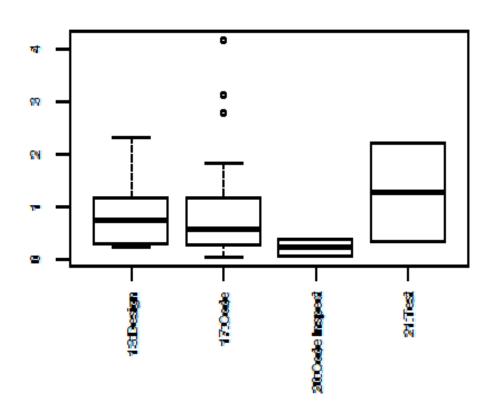
N	min	max	range
30	0.022	1.246	1.225

Median	Mean	SE
0.32222	0.394	0.0552

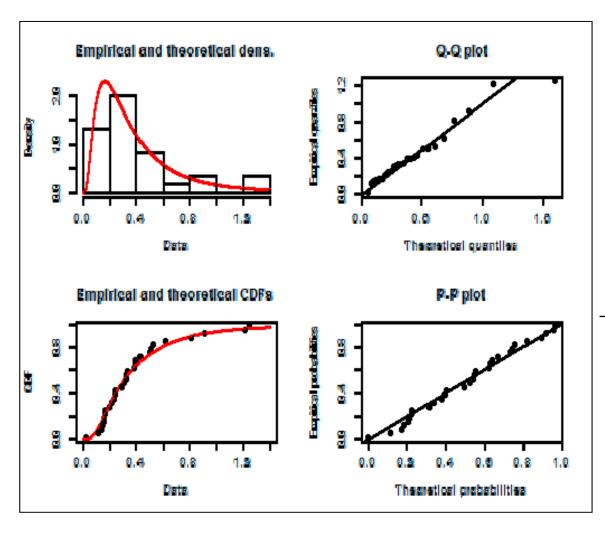
CI	Var	SD
0.113	0.0915	0.3025

例:あるプロジェクトのフェーズ毎のDIRの分布

boxplot of defect injection rate by phase in project_key 18



例: あるプロジェクトのDIRの対数正規分布への当てはまり



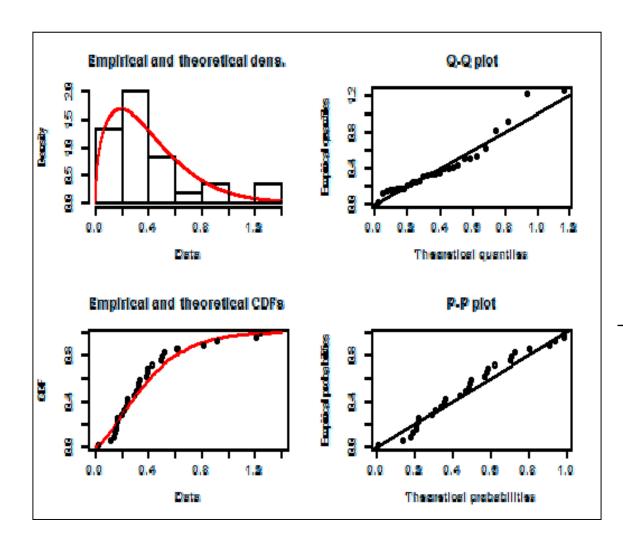
Maximum likelihood estimation

	Estimate	SD
meanlog	-1.1988	0.1434
sdlog	0.7852	0.1014

Anderson-Darling test

Anderson-Darling test	0.4221
p-value	0.8254

例: あるプロジェクトのDIRのワイブル分布への当てはまり



Maximum likelihood estimation

	Estimate	SD
shape	1.4331	0.1935
scale	0.4368	0.0589

Anderson-Darling test

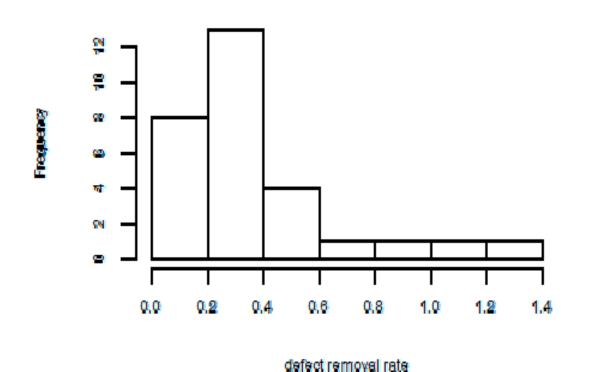
Anderson-Darling test	0.6421
p-value	0.6077

Defect removal rate

- Defect Removal Rate (以降、DRR)
 - = 不具合除去数/工数(人時)

例:あるプロジェクトのDRRの記述統計

histogram of defect removal rate in project_key 18

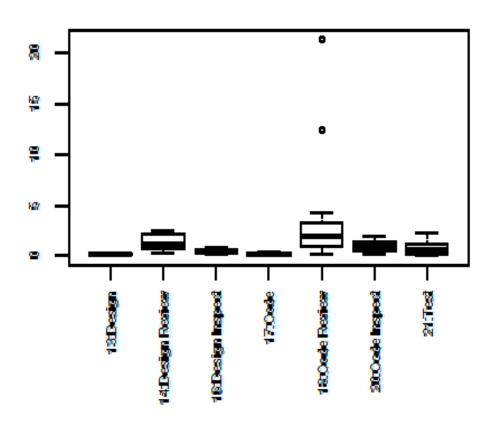


N	min	max	range
29	0.022	1.217	1.195

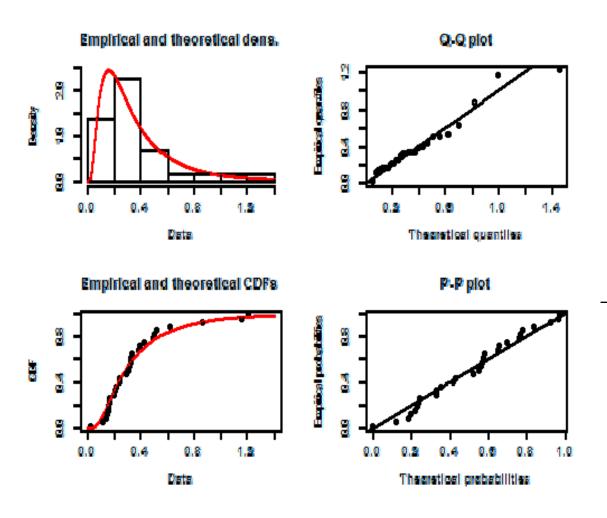
Median	Mean	SE
0.3165	0.371	0.0532

例: あるプロジェクトのフェーズ毎のDRRの分布

boxplot of defect removal rate by phase in project_key 18



例: あるプロジェクトのDRRの対数正規分布への当てはまり



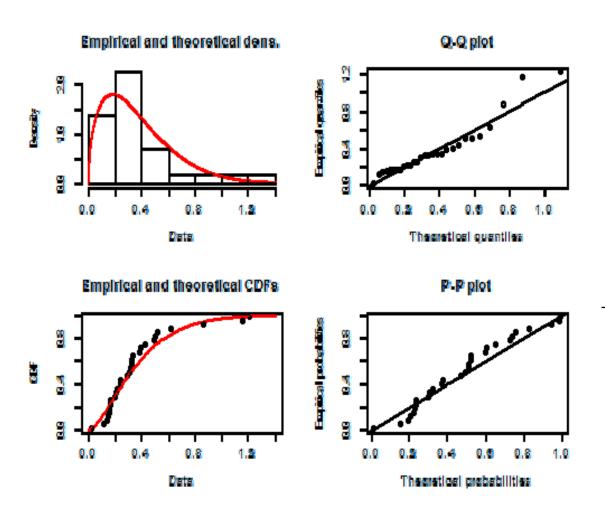
Maximum likelihood estimation

	Estimate	SD
meanlog	-1.2492	0.1427
sdlog	0.7686	0.1009

Anderson-Darling test

Anderson-Darling test	0.4956
p-value	0.7499

例: あるプロジェクトのDRRのワイブル分布への当てはまり



Maximum likelihood estimation

	Estimate	SD
shape	1.4425	0.1959
scale	0.4123	0.0562

Anderson-Darling test

Anderson-Darling test	0.7423
p-value	0.5232

DIR、DRRにおけるp値の分布

DIRにおけるp値の分布

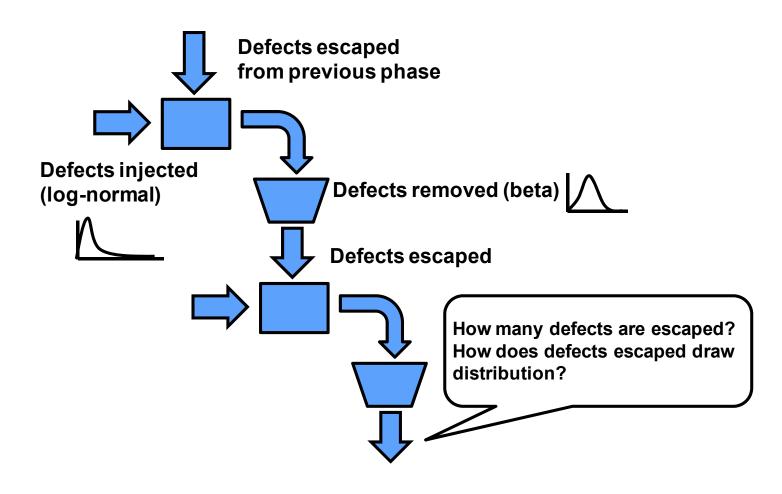
p値の範囲	SEMPR内のプロジェクトに占める割合
p値 > 0.7	約60%のプロジェクト(サイクル)
p值 > 0.9	約46%のプロジェクト(サイクル)
p值 > 0.95	約34%のプロジェクト(サイクル)

DRRにおけるp値の分布

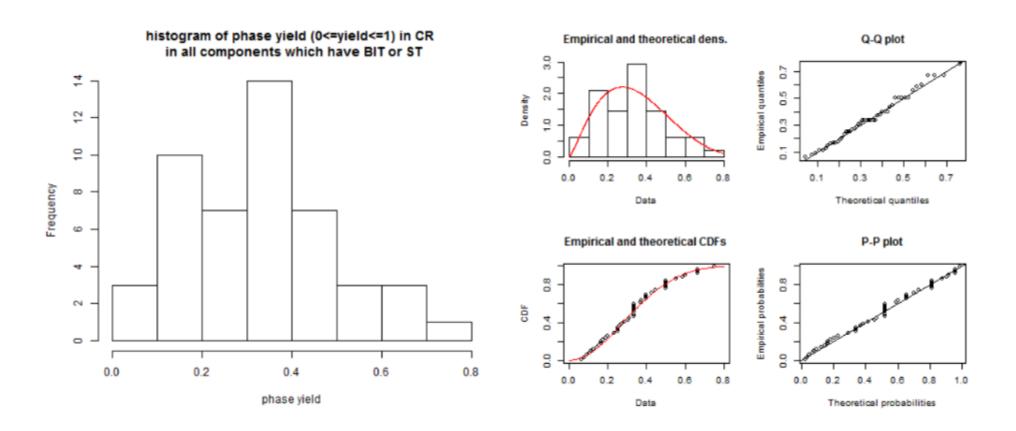
p値の範囲	SEMPR内のプロジェクトに占める割合
p値 > 0.7	約67%のプロジェクト(サイクル)
p値 > 0.9	約39%のプロジェクト(サイクル)
p値 > 0.95	約27%のプロジェクト(サイクル)

2.5 品質モデルの構築

●プロトタイプを構築



Phase Yieldの分布例(コードレビューフェーズ)



コードレビューにおけるPhase Yieldは、β分布に従う (他フェーズでは、上記のような傾向は発見できなかった)

目次

- 1. TSPとは
- 2. 研究のステップ
 - 2.1 研究の狙い
 - 2.2 TSPデータベースの構築
 - 2.3 データ品質評価
 - 2.4 データ分布調査
 - 2.5 品質モデルの構築
- 3. 研究のまとめとSEIの動向

研究のまとめ

- プロジェクト全体のデータ分布調査
- プロジェクト個別のDIR、DRRの分 布調査
- SEMPRを構築
- 現在、約250プロジェクトサイクル

データ分

品質モデ ルの構築

データ品

布調査

質評価

TSPデー タベース

研究 の構築 の狙い DIR、Phase Yield、各フェーズの 工数(計画)を活用した品質モデル のプロトタイプ構築(未評価)

- ツールの特性を考慮したデータ 品質モデルを構築
- 4つの属性について評価を実施

今後の展望

- プロジェクトのコンテキスト情報の蓄積
- 上流(要求分析 or 概念設計)から、ビルド/結合テストにおける不具合件数を予測する品質モデルの構築と評価

研究で感じたこと

• 分析の前のデータ品質評価

- データ収集のプロセスに着目し、モデルを構築すること
- やはり分析作業の多くの工数はデータクリーニング

• 色々と分析してみることが大事

- ただ、やみくもには行わない (仮説ベース)
- 仮説で悩む。ひたすら悩む

• R言語は便利

- 「グラフを出す」に対する工数は、相当減らせる
- 大事なこと(データ品質の考慮、仮説を立てる)に時間を振り向ける

• 社外発表は大事

- モチベーションの1つ
- 多くの多面的なフィードバックを得られる
- 本日も貴重な機会、ありがとうございました

連絡先

• 白井 保隆(シライ ヤスタカ)

E-mail: <u>yasutaka.shirai@toshiba.co.jp</u>

- TEL: 044-549-2405

- 株式会社 東芝 インダストリアルICTソリューション社 IoTテクノロジーセンター プロセス・品質技術開発部

ご質問等々ございましたら、どしどしご連絡ください。 定量データでソフトウェア・エンジニアリング領域を 一緒に発展させていきましょう!