

日本SPIコンソーシアム(JASPIC)

SPI Japan 2022

ITメトリクスを活用した外部ベンチマーキング事例の紹介 ～ プロセス改善効果を定量的に把握する手法 ～

2022年10月 6日

株式会社トイックス 佐藤 浩明



1. はじめに

2. メトリクスとは

3. メトリクス活用の期待効果

4. 外部ベンチマーキングの事例紹介(QCD編)

5. 外部ベンチマーキングの事例紹介(追跡調査編)

6. おわりに

● 株式会社トイックス (TOiNX)

- 情報システムの企画・コンサルティングから開発，運用，保守そして情報機器販売・リースに至るまで，総合的な情報システムサービスを提供しています。
- 従業員数：722名(2022年4月時)
- 本社所在地：宮城県仙台市
- URL：<http://www.toinx.co.jp/>

● 担当業務

- コンサルティング・サービス
 - ◆ ITメトリクスの導入・解析支援
 - ◆ 標準プロセスの整備・展開支援
 - ◆ 各種ツールの開発



【参考】トイックスにおけるSPI活動の経緯

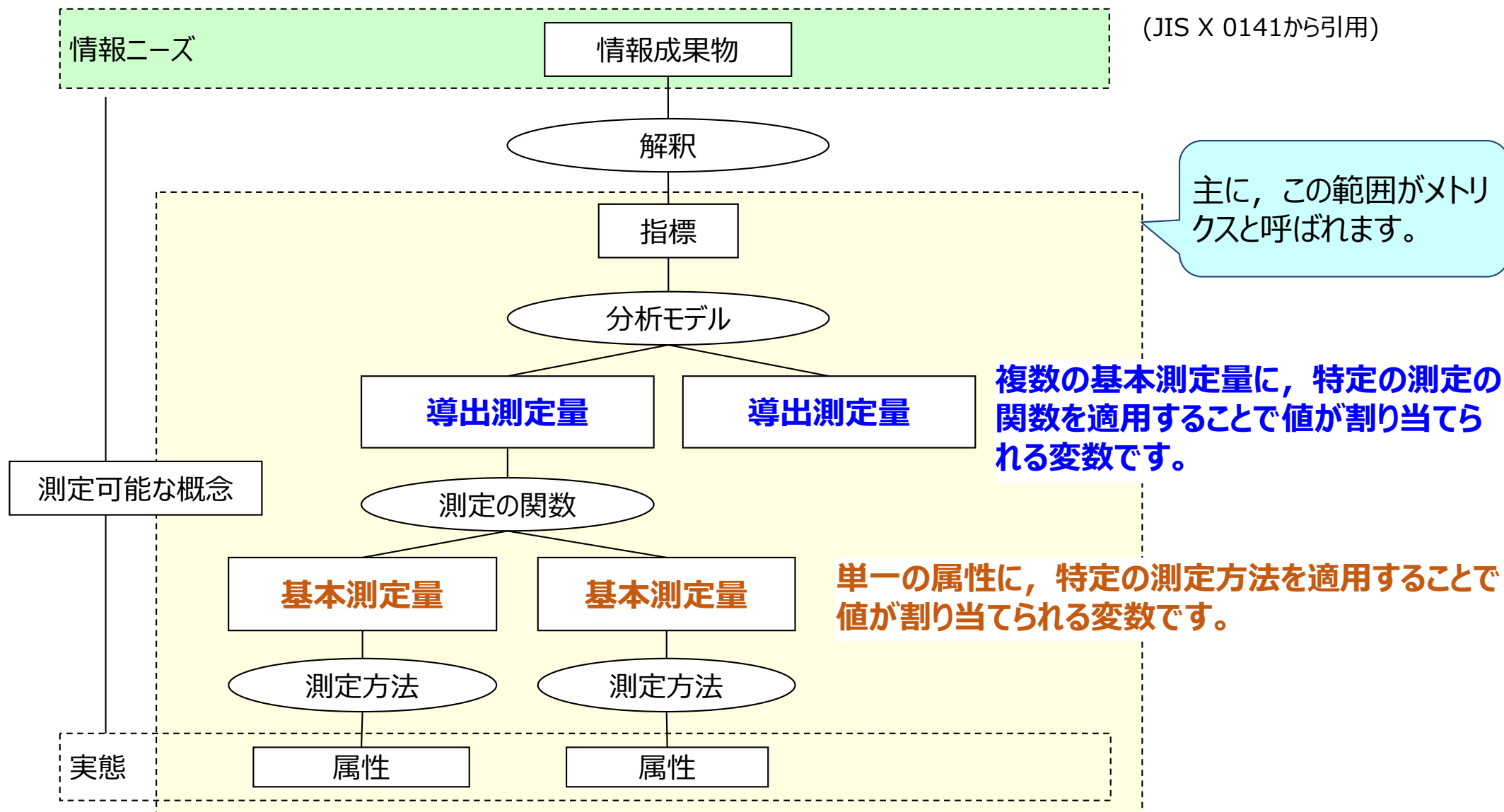
年度	主な活動
2001	SW-CMM®に関する動向調査・SPI推進計画の策定。
2002	SPI担当部門の設置。システム開発プロセス標準の整備と品質保証活動の開始。 第1回目の正式アプレイザル。Lv2とのギャップ分析。
2003	第2回目の正式アプレイザル。Lv2の達成を確認。
2005	第3回目の正式アプレイザル。Lv3とのギャップ分析。開発部門から独立したPMOの設置。
2007	第4回目の正式アプレイザル。Lv2の達成を確認。
2008	第5回目の正式アプレイザル。Lv3の達成を確認。
2010 ～ 2011	定量的プロジェクトマネジメントの実践を目指し「メトリクス分析支援ツール」を開発。(プロジェクトの実績データが、ある程度蓄積できたため)
	プロジェクトの立上げから完了までの計画・実績情報を一元管理・可視化する「マネジメント支援システム(Co.守り)」を開発。
	第6回目の正式アプレイザル。Lv3の達成を確認。(以降、自社での評価に切り替え。)
2012	メトリクス解析結果 の公開を開始。以降、毎年、最新版を公開。 併せて、各種のメトリクス応用ツールを順次開発し提供。
2016 ～	自社内でのSPI活動をベースにしたコンサルティング・サービス「ITメトリクス導入・解析支援サービス」を展開。

メトリクスの公開と活用が、SPI活動定着のターニングポイントでした。

- 開発プロジェクトの成功には、自社の実績データを解析したメトリクスを用いた、定量的マネジメントの実践が有効です。
- 一方で、メトリクスの活用や解析には多くの検討事項があります。
- トインクスでは、2016年度から「ITメトリクス導入・解析支援サービス」を展開し、多くのお客さまを支援しています。
- 今回は、このサービスの中でもニーズが高い、メトリクスを活用した、外部ベンチマーキングの事例を紹介します。

2. メトリクスとは

2.1. メトリクスの定義



本発表で扱うメトリクスは、主に「プロセス実績」に関するものです。

導出測定量(Z)

結合テスト密度(FP) とは

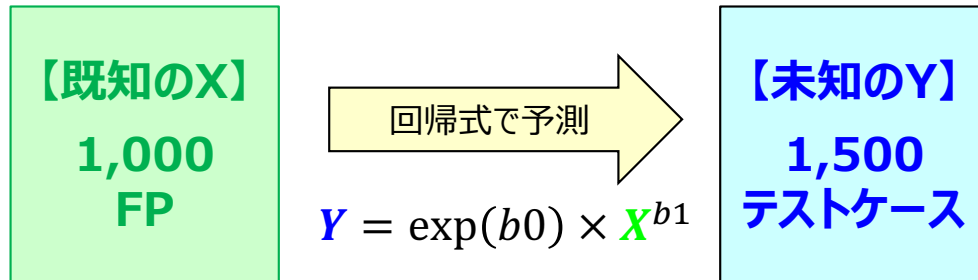
例えば...

結合テストケース数 と FP の関係

基本測定量(Y)

基本測定量(X)

① 予測・妥当性評価

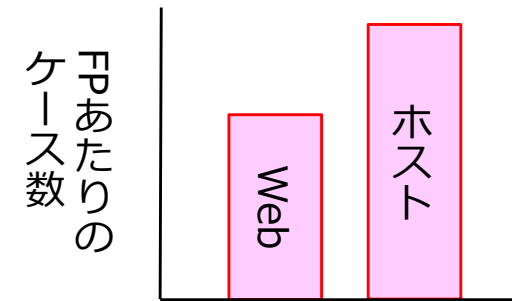


回帰式を利用して、「既知のX」から「未知のY」を予測します。

1,000FPのプロジェクトでは、1,500ケース程度を設定するようだ！！



② 傾向分析



サンプルごとの「 $Z(=Y \div X)$ 」の代表値を算出して、傾向を把握します。

どうやら、ホストの方がテストケースを多く設定する傾向のようだ！！



(注) 予測・妥当性評価には、上記の単回帰のほかに、重回帰を用いることもあります。

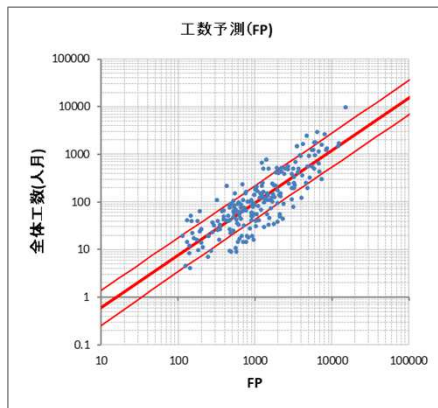
また、通常は確率変数は大文字、観測値は小文字で表しますが、本資料では特に区別せず大文字で記載します。

2.3. プロセス実績モデルとプロセス実績ベースライン

8

● プロセス実績モデル (PPM : Process Performance Model)

- 開発プロジェクトのさまざまな場面で利用する予測や妥当性評価の仕組み



「既知のX」から「未知のY」を予測

例1：開発工数から最適な工期を予測

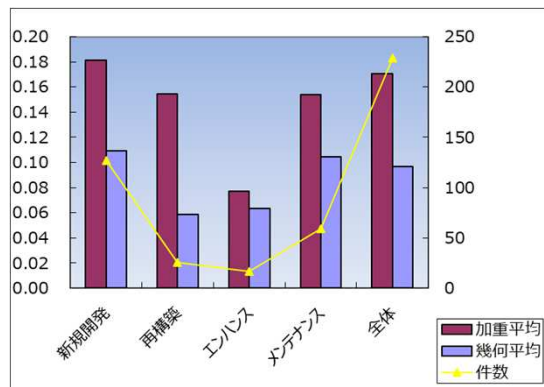
例2：開発規模から最適なテストケース数を予測

左図の赤線(中央)で示す回帰線を求める計算式自体が PPMとなります。



● プロセス実績ベースライン (PPB : Process Performance Baseline)

- 開発組織の代表的な姿を現す基準値と、上下限値の3点セット



PX_ESTIMEFP

No	集計区分	件数	加重平均	幾何平均	標準偏差	-3σ点	-2σ点	-σ点	+σ点	+2σ点	+3σ点
1	新規開発	127	0.181	0.109	2.377	0.008	0.019	0.046	0.259	0.616	1.465
2	再構築	26	0.154	0.058	3.616	0.001	0.004	0.016	0.211	0.763	2.759
3	エンハンス	17	0.077	0.063	2.205	0.006	0.013	0.029	0.140	0.308	0.679
4	メンテナンス	59	0.154	0.105	2.107	0.011	0.024	0.050	0.220	0.464	0.978
全体		229	0.170	0.097	2.491	0.006	0.016	0.039	0.240	0.599	1.491

ITメトリクスでは、幾何平均値 $\pm 1\sigma$ (全データの約2/3を含む範囲) を利用することが多いです。



2.4. プロセス実績モデル(PPM)とは-1

- ある変数の値で、他の変数の値を予測します。
 - 見積もりや、見積もり結果の妥当性評価に利用します。
- べき乗モデルを用いて、プロジェクトの規模に応じた曲線で予測します。

(例) **工数(人月)** = $\exp(b0) \times \text{FP}^{b1}$... 「b0」と「b1」は回帰分析で求めた係数

基本測定量(Y)

基本測定量(X)

単回帰分析(対数)

【べき乗予測】

信頼率 ±1σ

X(入力)	1000
Y(予測)	109.9
Y(下限)	48.5
Y(上限)	249.2

①

②

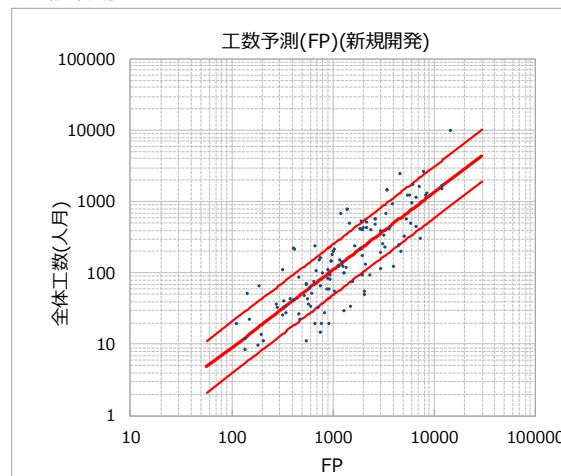
【単回帰分析】 モデル: $y = \exp(b0) * x^{b1}$

	b0	b1	$\exp(b0)$
回帰係数	-2.8191974	1.0884605	0.0596538
標準誤差	0.4853579	0.0669234	1.6247564
t 値	-5.808492	16.264285	
P 値	5.204E-08	3.105E-32	

【寄与率】

R2	68.6%
R*2	68.4%
R**2	68.1%

【両対数散布図】



①にFP値を入力すると、②に工数の予測値と上下限値を表示します。

見積もりや、見積もり結果の妥当性評価に利用します。

散布図の両軸を対数にしているため直線に見えますが、曲線の予測となります。(次スライドで説明します。)

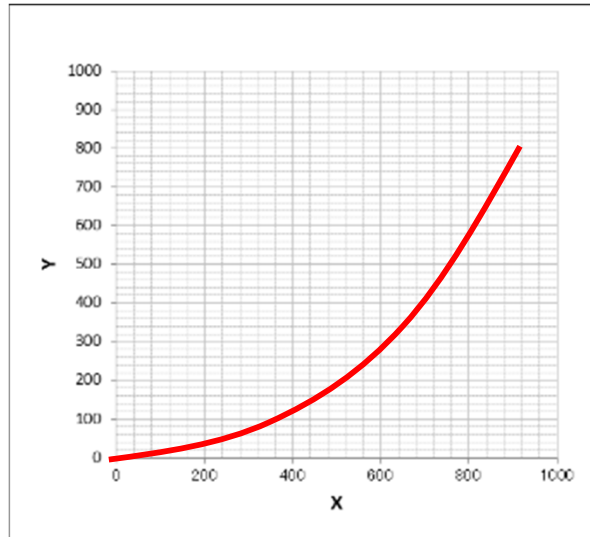


毎年、最新の実績データで約130種類のPPMをプロジェクトに提供。

2.4. プロセス実績モデル(PPM)とは-2

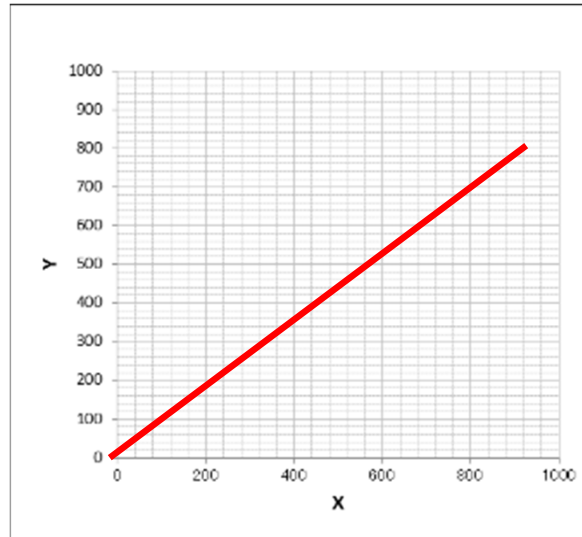
- べき乗モデルは，回帰係数「 b_1 」の値で形状が変化します。

パターンA ($b_1 > 1$)



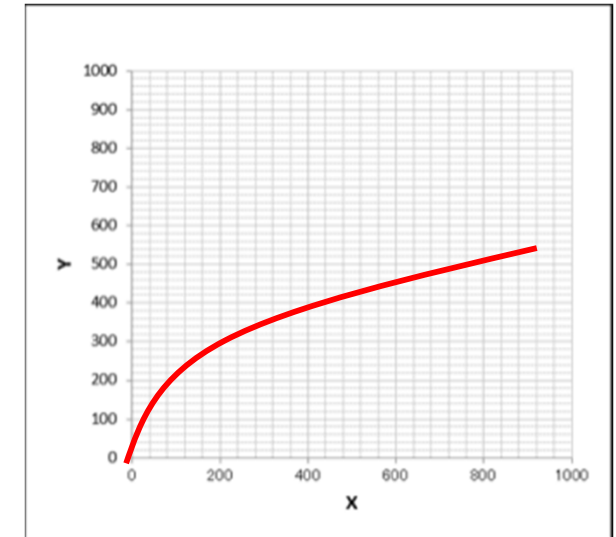
Xが大きくなるにつれ，
Yが増加していく。

パターンB ($b_1 = 1$)



XとYは**比例関係**であり，
直線となる。

パターンC ($b_1 < 1$)



Xが大きくなるにつれ，
Yの増加が減少する。

- 層別が不十分な場合，真の関係が「パターンA」でも「パターンC」になることがあります(バラつきが大きいと傾きが寝る)。

2.5. プロセス実績ベースライン(PPB)とは

11

- 1単位あたりの代表的な値と、その上下限值を表すものです。
 - 例えば、「FP生産性」の定義は以下のとおりです。

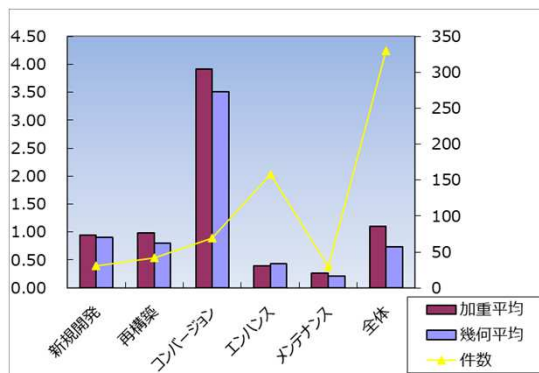
$$(例) \text{ FP生産性} = \text{FP} \div \text{工数(人月)}$$

導出測定量(Z)

基本測定量(Y)

基本測定量(X)

- 通常、層別ごとの傾向比較や、年度推移の把握に利用します。
- 「Z」の値で、暫定的に予測・妥当性評価をする場合は**要注意**です。
 - ITメトリクスの多くは、前述のパターンA or Cのため、予測に乖離が生じます。



製造業などと違い、バラツキの大きいITメトリクスでは、幾何平均値 $\pm 1\sigma$ の利用を推奨します。

PX_PRODLOC

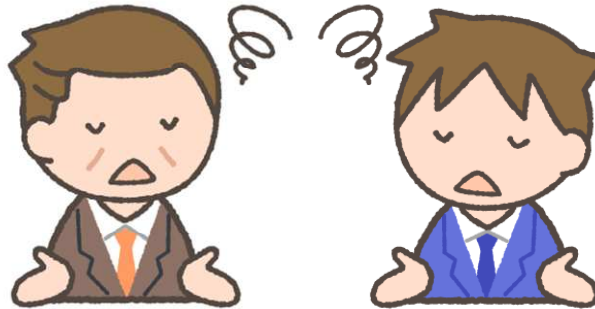
No	集計区分	件数	加重平均	幾何平均	標準偏差	-3σ点	-2σ点	-σ点	+σ点	+2σ点	+3σ点
1	新規開発	31	0.950	0.901	1.697	0.185	0.313	0.531	1.529	2.594	4.401
2	再構築	42	0.978	0.805	1.755	0.149	0.261	0.459	1.412	2.479	4.350
3	コンバージョン	69	3.914	3.513	2.610	0.198	0.516	1.346	9.169	23.930	62.455
4	エンハンス	158	0.400	0.439	2.139	0.045	0.096	0.205	0.940	2.010	4.300
5	メンテナンス	30	0.268	0.214	1.876	0.032	0.061	0.114	0.402	0.754	1.414
	全体	330	1.096	0.734	3.196	0.023	0.072	0.230	2.347	7.502	23.976



3. メトリクス活用の期待効果

メトリクスの効果とは??

メトリクスを使うと
何ができるの？



利用すると, どんな
メリットがあるの？

メトリクスを用いた「定量的マネジメント」の期待効果について
解説します。

測れないものはマネジメントできない

- 開発方法論やリスク管理で著名なトム・デマルコ氏や、経営学者のピーター・ドラッカー氏らが繰り返し唱えた言葉です。
- プロジェクトの今後の見通し(予測)を立て、現状を正確に把握し制御するためには、カンと経験だけではなく、定量的マネジメントが求められます。
- 言い換えれば・・・

測れるものは管理・改善できる!!

3.3. 定量的マネジメントへの期待-1

15

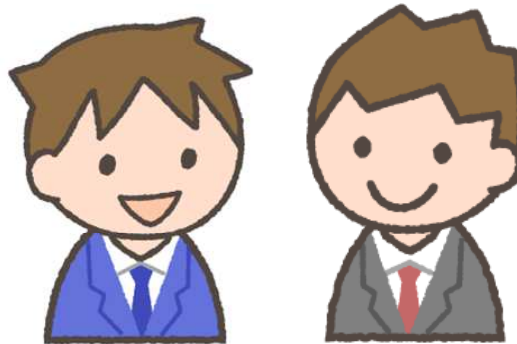
● 見積もり精度の向上と計画内容の妥当性評価

- 工期や工数見積もりの一手法として活用
- プロジェクト計画の客観的な妥当性評価

本PJの計画工数は「100人月」です。
今回の開発期間は「10ヶ月」と見積もりました。

【注意】

妥当性評価では、積算法や類推法の見積もり結果を、マトリクスを用いた「パラメトリック法」で評価するのが現実的です。



PM

上位責任者

本当に大丈夫???
若干短いような気がするのだけど...

過去の実績から、約68%の確率で「11.8～15.1ヶ月」が妥当と評価します。
作業の組み立てや、依存関係に無理がないことを再点検してください。

3.3. 定量的マネジメントへの期待-2

● 定量的品質管理の実践

- レビューやテストなど品質に関する基準値の設定
- 基準値との比較による定量的な監視・制御

本PJの開発規模は「1,000FP」です。
結合テストでは「500テストケース」を設定します。

【注意】
工程進行中の品質管理には、SRGMなどのメトリクス以外のツールも必要になります。



品質の確保は大丈夫???
少ないような気もするのですが...

過去の実績から、同規模のPJでは「750～1,500テストケース」を設定しています。
前工程の検出欠陥数も基準値の上限付近だったので、テストケースを追加してください。

● 開発組織の実力値の把握

- 品質，生産性，工期に関する実力を定量的に把握
- 特定した改善機会に実現可能なプロセス実績目標を設定

トインクスの生産性は、
同業他社と比較してどうなの
かね。

【注意】

FP生産性では，開発プロジェクトの生産性を
正確に把握できないことがあります。

詳細は，ユニシス技報(2016年9月発刊
Vol.36 No.2 通巻129号)の『ソフトウェア開
発プロジェクトの生産性評価に関する事例』を
参照してください。



お客さま，
経営者



SPI担当者

感覚的にはそれほど低くないとは思いますが...

組織全体のFP生産性は「〇〇/
人月」です。外部ベンチマークよりも
高いです。
ただし，A分門は「△△/人月」と
若干低い状態ですので，今年度
は「□□/人月」までの向上を目標
とします。

3.4. 期待効果のまとめ

【期待効果1】

見積もり精度の向上と
計画内容の妥当性評価

- 工期や工数の予測・評価に利用するメトリクスを確立

- 見積もり時や、見積もり結果の妥当性評価に活用

プロジェクト計画に対する早期段階での妥当性評価

(リスクを特定・対応)

【期待効果2】

定量的品質管理の実践
による成果物品質の向上

- レビューやテストに関するメトリクスを確立

- テストケースや指摘・検出欠陥数の目標設定と、品質の判断材料として活用

全工程の品質目標値の設定と監視が可能

(上流から計画的に品質を作り込む)

【期待効果3】

開発組織の実力値の把握

- 生産性、品質、工期などに関する組織の実力値を定量的に特定

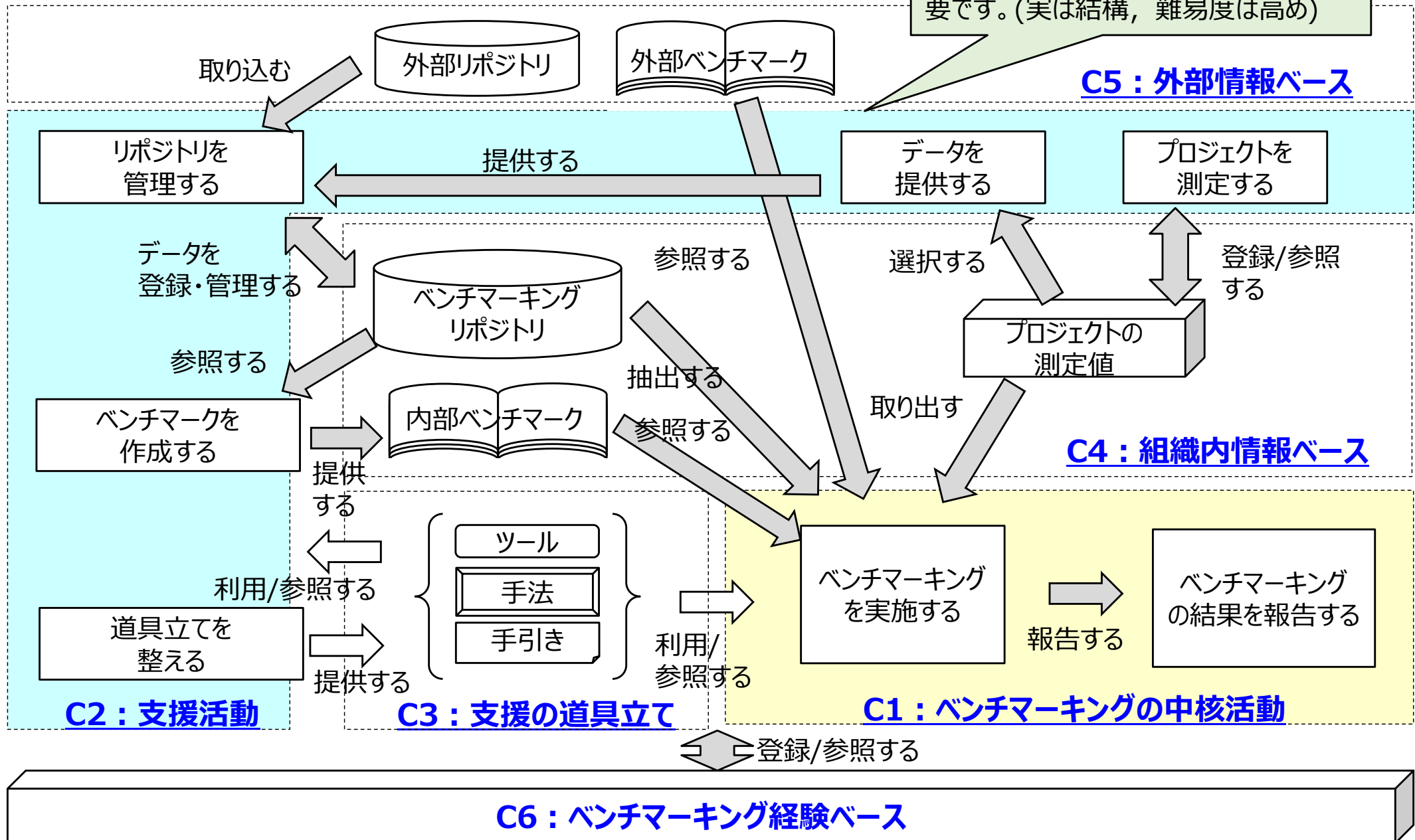
外部・内部ベンチマーキングで、自社・組織別の状況を把握

(実現性のある改善目標を設定)

【参考】ベンチマーキングの枠組みモデル

19

ベンチマーキングの成功には、中核活動だけではなく、水色で示す支援活動が重要です。(実は結構、難易度は高め)



4. 外部ベンチマーキングの事例紹介(QCD編)

4.1. 事例紹介の前に

- 以降は、トインクスが実施した外部ベンチマーキングの事例(一部)を紹介します。
- A社を対象とした「外部ベンチマーキング結果報告書」の抜粋ですが、メッセージ性を変えない程度にデータを加工しています。
- なお、本事例では、実力値を最も反映する**新規開発**を対象に、**品質・コスト・工期**を評価しています。
- 外部ベンチマークの**全業種**と、A社が該当する**金融・保険業**を比較の対象としています。

(参考)

- ベンチマーキングとは
 - ◆ 自社(または組織)の実績データと、業界標準団体などが提示する外部ベンチマークを比較するプロセスです。
 - ◆ 典型的には、業界における位置付けを明確にし、プロセスの強みや、改善機会を特定するために定期的に実施します。

4.2. 利用データの特徴

- 本ベンチマーキングでは、以下のデータを利用しました。

データ区分	データ数	取得元	備考
A社データ	70~79件	2004から2020年度の「新規開発」の実績データを利用しました。	・プロジェクトの失敗・成功を問わず、該当するデータをすべて利用しています。 ・グラフデータに対応するメトリクスの解析時は、一定のルールで外れ値を除外しています。
外部ベンチマーク	①全業種 90~303件 ②金融・保険業 84~218件	IPA(※1)が公開する以下のデータを利用しました。 ①「ソフトウェア開発データ白書2018-2019」の新規開発のグラフデータ(Copyright 2018 IPA) ②「ソフトウェア開発データ白書2016-2017 金融・保険業編」の新規開発のグラフデータ(Copyright 2016 IPA)	・グラフデータの解析時は、A社と同じルールで外れ値を除外しています。 ・グラフデータがない場合、データ白書が提示する中央値または平均値を利用します。 ・金融・保険業編は2018-2019のグラフデータ数が少ないため、2016-2017を利用しています。

- 外部ベンチマークは、データ白書が提示する解析結果ではなく、**グラフデータを解析した結果を利用**します。(理由は次スライドで)
- また、外部ベンチマークとの整合性を保つため、工数と期間は、開発5工程(※2)のデータを利用しています。

※1 独立行政法人情報処理推進機構(Information-technology Promotion Agency : IPA)

※2 基本設計～総合テストの5工程を指す。

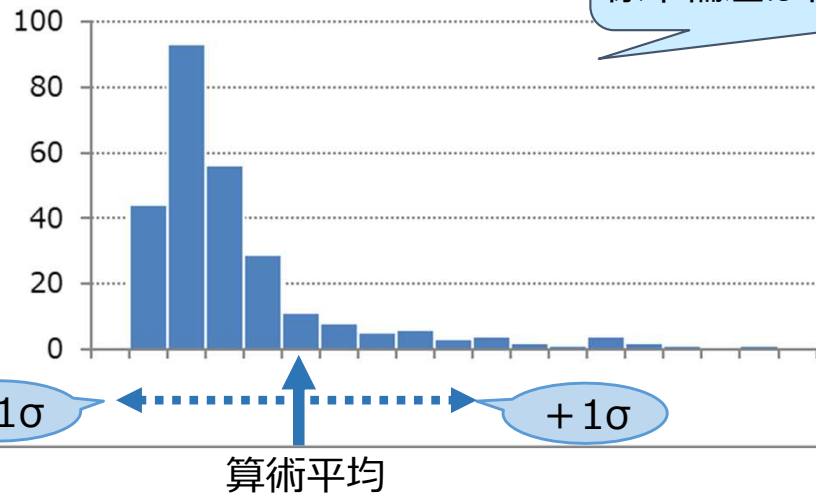
【参考】なぜグラフデータを解析するのか？

- ITメトリクスは、多くの場合、歪んだ分布のため、変数変換で正規分布として扱います。

【FP生産性のヒストグラム】

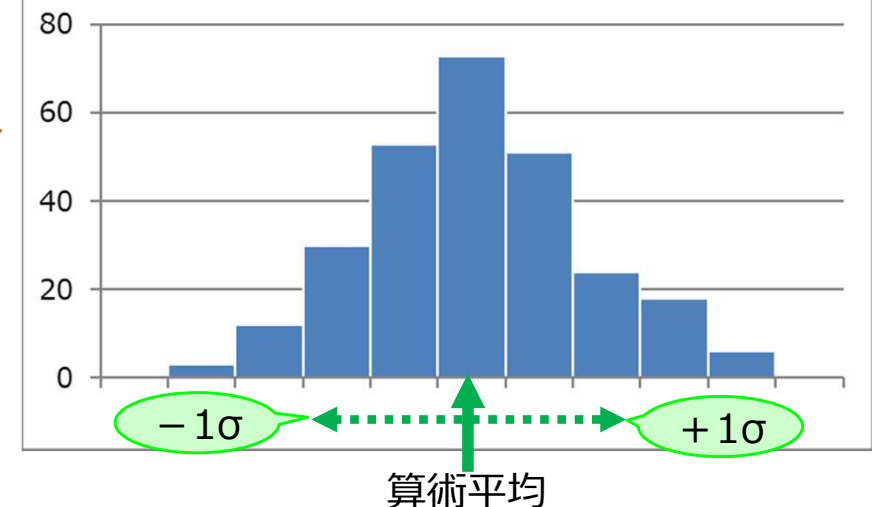
FP生産性(真数)

非正規分布の場合、算術平均や標準偏差は利用できません。



正規分布の性質を利用して算術平均や標準偏差を算出します。

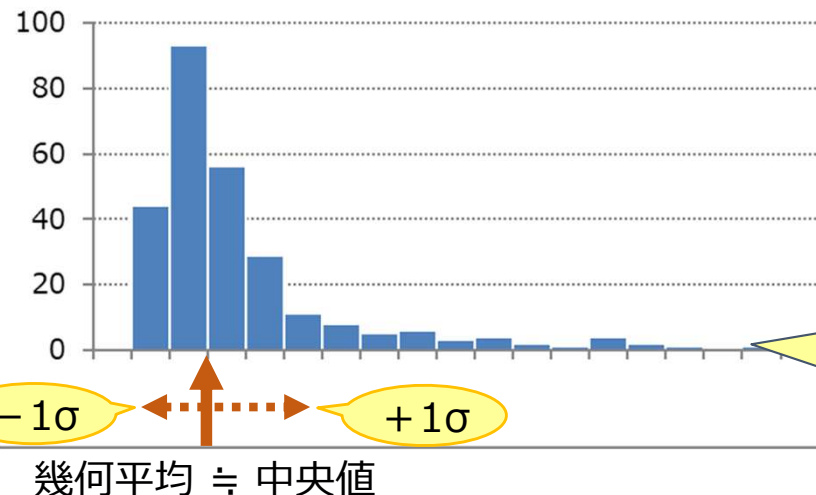
FP生産性(変数変換後)



対数変換

FP生産性(真数)

指数変換

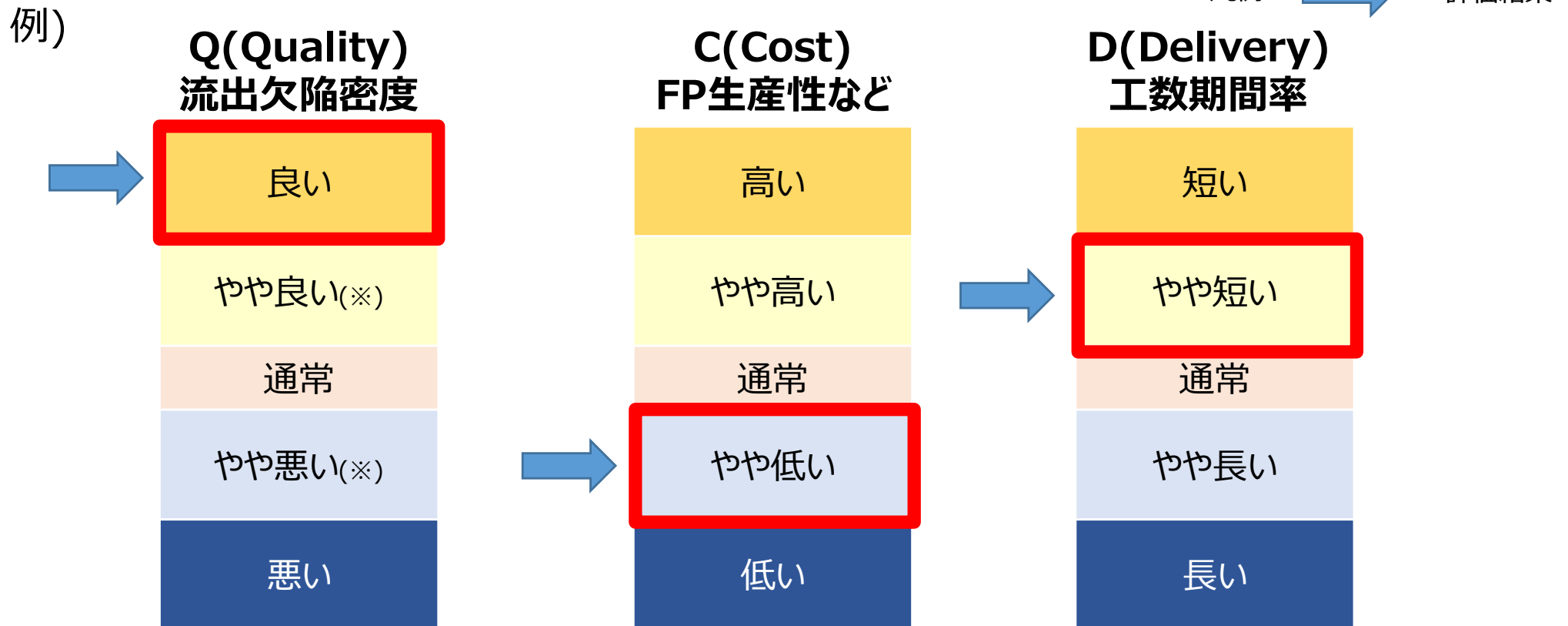


後述する「t検定」は、正規性が前提になります。このため、変数変換後の値を用いて検定することになります。

対数変換後の算術平均を真数に戻したものを「幾何平均」と呼び、集団の中心的傾向を示す代表値になります。なお、 $\pm 1\sigma$ も左右で大きさが変わります。

4.4. 評価結果の提示方法(例)

- トインクスの外部ベンチマーキングでは、QCDを5段階で評価します。



上の例では、QCDに対する評価は以下のとおりです。

- ・ 品質(Q)は「良い」 : 高品質傾向
- ・ 生産性(C)は「やや低い」 : 低生産傾向の可能性がある
- ・ 工期(D)は「やや短い」 : 短工期傾向の可能性がある

※ 上記の「やや良い」は、良い可能性があるという意味です。同様に「やや悪い」とは悪い可能性があるという結果を指します。
(**良い・悪いの程度ではなく、可能性の確かさ**と捉えてください。)

4.5. 5段階の評定基準

- QCDの定量的な評定基準は、以下のとおりです。

評価結果	評定基準
良い	①外部ベンチマークよりも良い結果で、t検定が有意水準5%で有意 ②べき乗回帰式の重心での傾向比較で、外部ベンチマークよりも20%以上良い結果
やや良い	①外部ベンチマークよりも良い結果で、t検定が有意水準20%で有意 ②べき乗回帰式の重心での傾向比較で、外部ベンチマークよりも10%以上20%未満の良い結果
通常	①または②で、外部ベンチマークとほぼ同等の結果
やや悪い	①外部ベンチマークよりも悪い結果で、t検定が有意水準20%で有意 ②べき乗回帰式の重心での傾向比較で、外部ベンチマークよりも10%以上20%未満の悪い結果
悪い	①外部ベンチマークよりも悪い結果で、t検定が有意水準5%で有意 ②べき乗回帰式の重心での傾向比較で、外部ベンチマークよりも20%以上悪い結果

※ 以降は、上記の①を「PPB比較」、②を「PPM比較」と表記します。

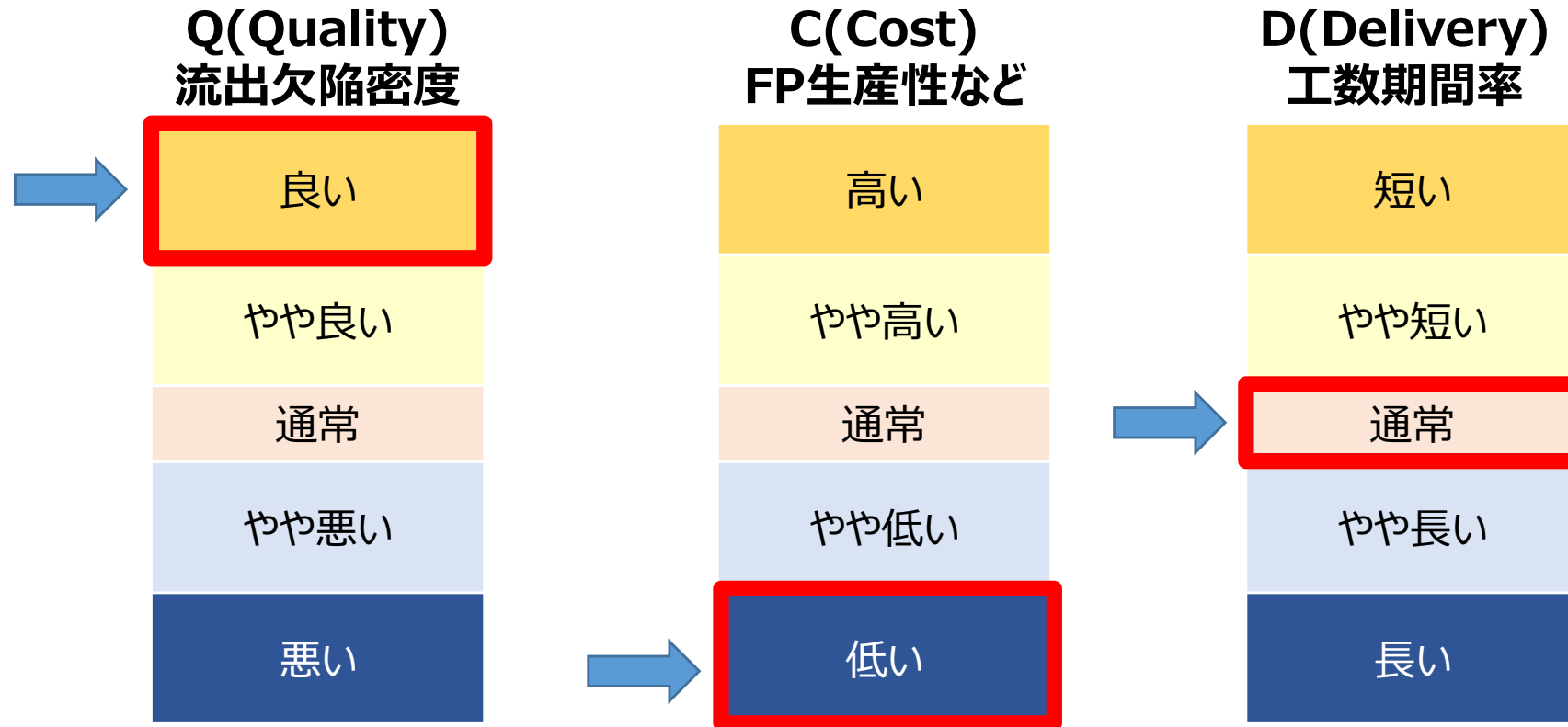
4.6. 評価対象メトリクス

- QCDの評価には、以下のメトリクスを用いました。

区分	メトリクス(Z)		基本測定量(Y, X)	評価方法	概要
品質 (Quality)	1	流出欠陥密度 (FP)	Y : 流出欠陥数 X : KFP	PPB比較	・ 1,000FPあたりの流出欠陥数を表します。 ・ システムの最終品質を表す指標として国内・国外で用いられます。
生産性 (Cost)	2	FP生産性	Y : FP X : 5工程工数(人月)	PPB比較	・ 1人月あたりに製造したFP量を表します。 ・ 主に、国内の生産性指標として用いられます。
	3	工数予測 (FP)	Y : 5工程工数(人月) X : FP	PPM比較	・ 1FPあたりの製造に費やした工数を表します。 ・ 主に、国外の生産性指標として用いられます。
工期 (Delivery)	4	工数期間率	Y : 5工程期間(月数) X : 5工程工数(人月)	PPM比較	・ 1人月あたりに要する開発期間を表します。 ・ 工期の長短を表す指標として国内・国外で用いられます。

4.7. ベンチマーキング結果(vs 全業種)

- 全業種と比較したQCDの評価は以下のとおりです。



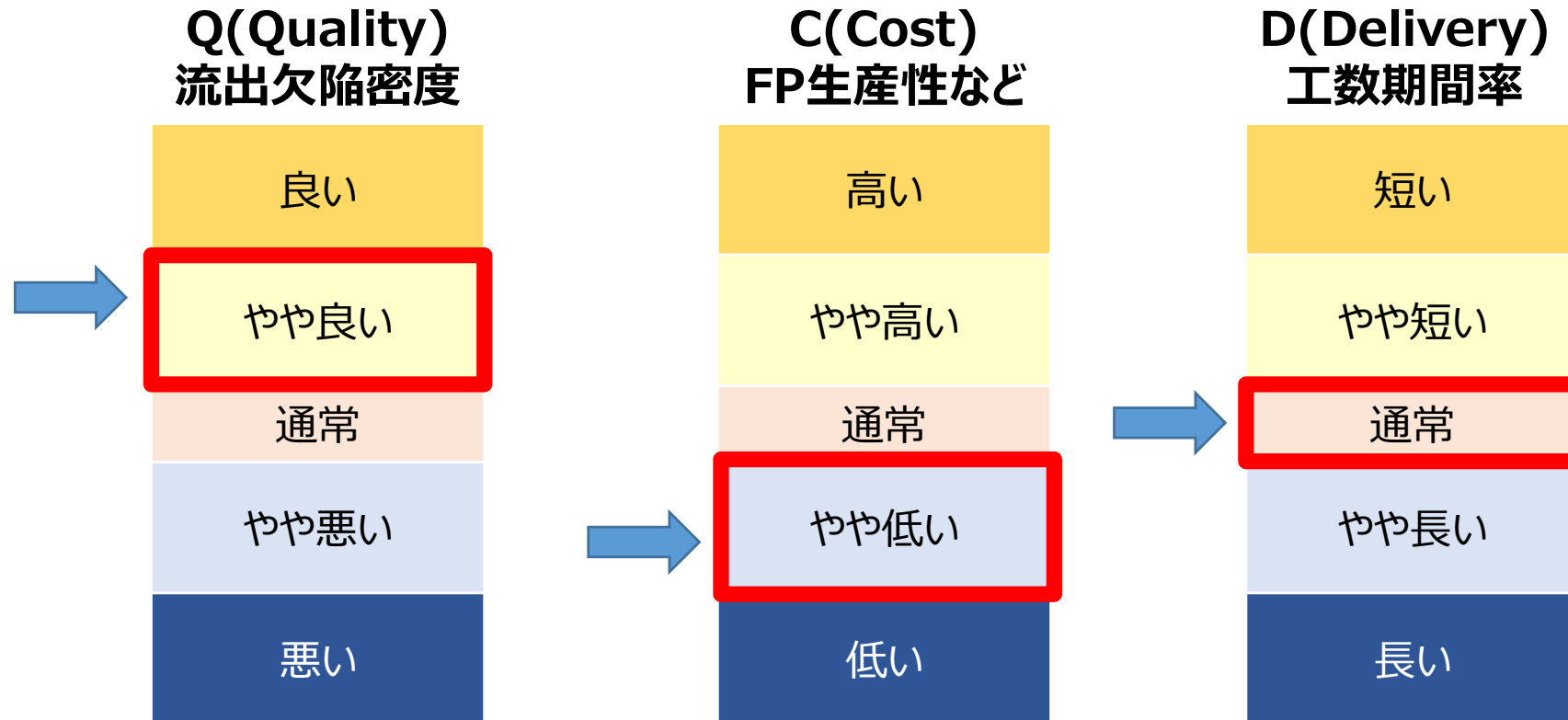
QCDは、以下の評価結果となりました。

- ・ 品質(Q)は「良い」
- ・ 生産性(C)は「低い」
- ・ 工期(D)は「通常」

高品質ですが、生産性に改善の余地があります。

4.8. ベンチマーキング結果(vs 金融・保険業)

- 金融・保険業と比較したQCDの評価は以下のとおりです。



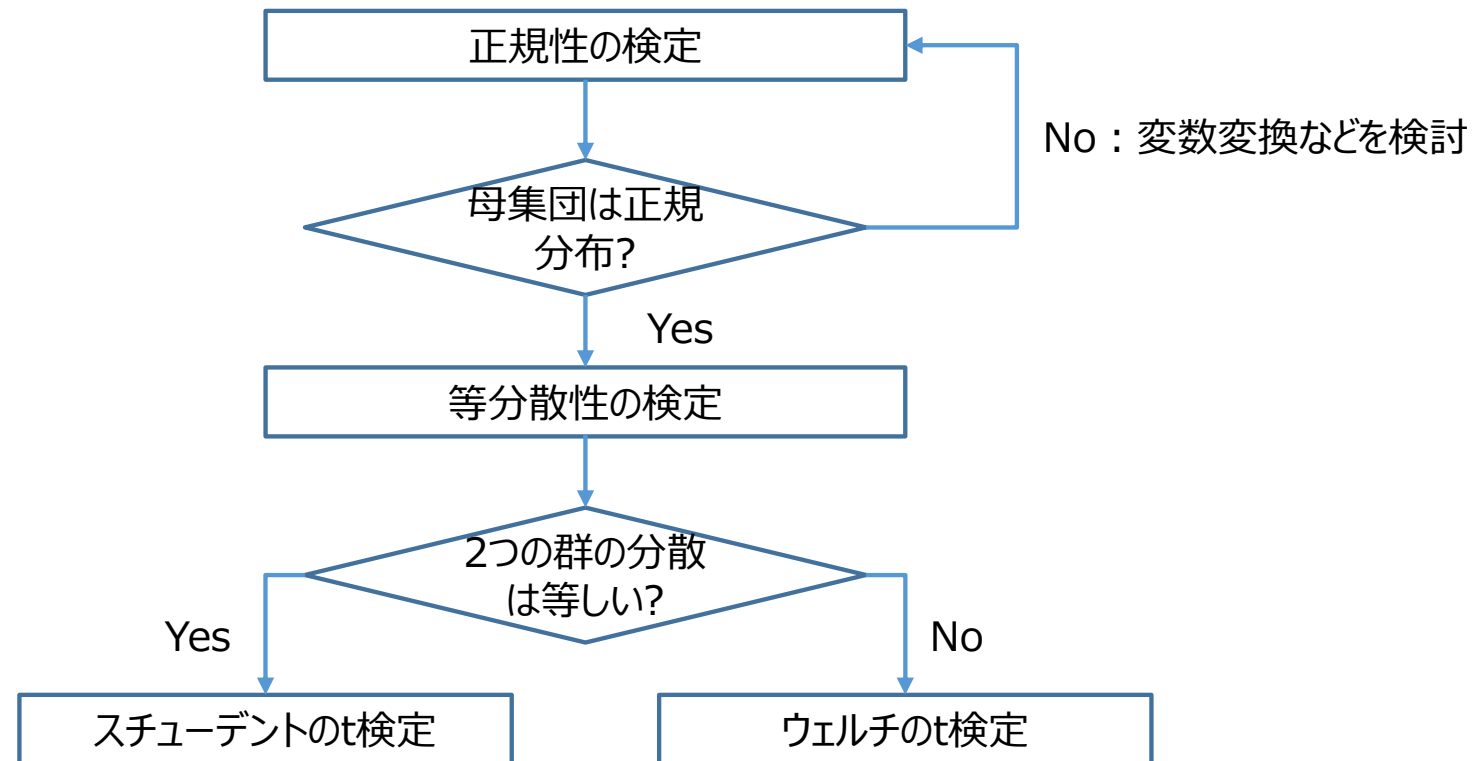
QCDは、以下の評価結果となりました。

- 品質(Q)は「**やや良い**」
- 生産性(C)は「**やや低い**」
- 工期(D)は「**通常**」

やや高品質ですが、生産性に改善の余地がある可能性があります。

【参考】データに対応がない2標本のt検定

- 次スライド以降に評価の詳細を提示しますが、その前に。



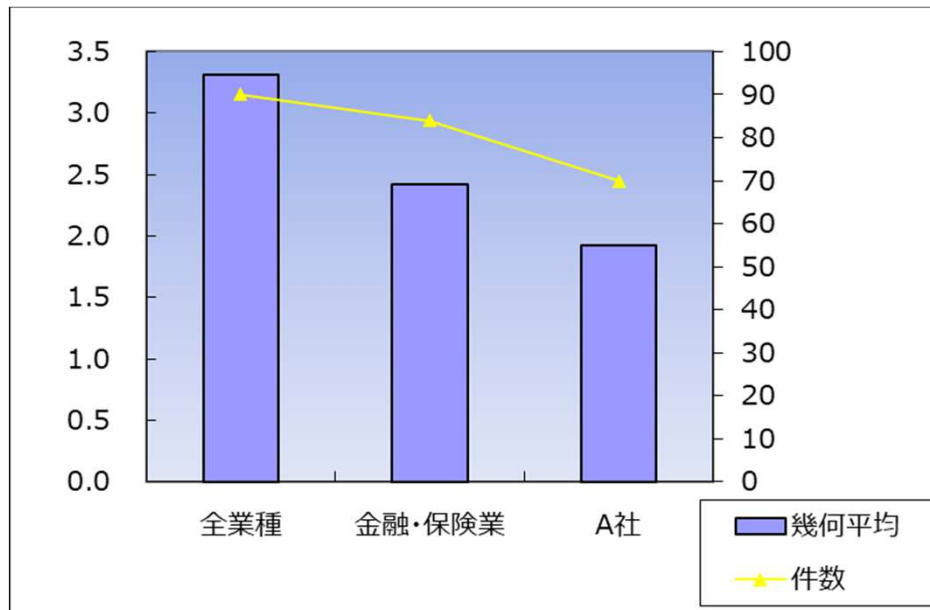
- 最初から等分散性の考慮が必要のないウェルチを用いればよいという説も。。。
 - ただし、ウェルチのt検定では自由度を2群の分散で補正します。補正された自由度は本来の自由度($n_A + n_B - 2$)以下になり、検定結果は有意になりにくくなります。

4.9. 品質「流出欠陥密度(FP)」のPPB比較

30

流出欠陥密度(FP)

No	集計区分	件数	幾何平均	標準偏差
1	全業種	90	3.309	1.012
2	金融・保険業	84	2.418	1.005
3	A社	70	1.925	1.003



1. 傾向分析

A社の1,000FPあたりに流出した欠陥数(幾何平均)は、**1.93件**です。

- ・全業種の**0.58倍**
- ・金融・保険業の**0.80倍**

であり、値は両ベンチマークよりも低いです。

2. t検定結果(vs 全業種)

P値は極めて小さく、有意水準5%で有意でした。このため、A社の品質は「**高い**」と評価しました。

3. t検定結果(vs 金融・保険業)

P値は**10.3%**で、有意水準20%で有意でした。このため、A社の品質は「**やや高い**」と評価しました。

比較対象	件数	平均	標準偏差	差(A-B)のt検定 (Welch法)							P(両側)
				差(A-B)	自由度	合成分散	合成標準偏差	to値	有意水準	t値	
LN(A社)	70	-6.2527	0.72968								
LN(IPA全業種))	90	-5.7112	1.10037	-0.54159	154.4187	0.02106	0.145119	-3.73203	5%	1.975488	0.027%

帰無仮説を棄却する

比較対象	件数	平均	標準偏差	差(A-B)のt検定 (Welch法)							P(両側)
				差(A-B)	自由度	合成分散	合成標準偏差	to値	有意水準	t値	
LN(A社)	70	-6.2527	0.72968								
LN(IPA(金融・保険業))	84	-6.0249	0.98961	-0.22785	149.8862	0.019265	0.138798	-1.64158	20%	1.287259	10.28%

帰無仮説を棄却する

全業種：「グラフデータ 図表9-2-2 ● FP 発生不具合密度の基本統計量(新規開発)」を解析して利用

金融・保険業：「グラフデータ 図表9-2-4 ● FP 発生不具合密度の基本統計量(新規開発, IFPUG グループ)」を解析して利用

※ レイアウト上、3標本の結果を同時に掲載していますが、**全業種と金融・保険に対する評価は独立**です。(検定の多重性はなしとします。)

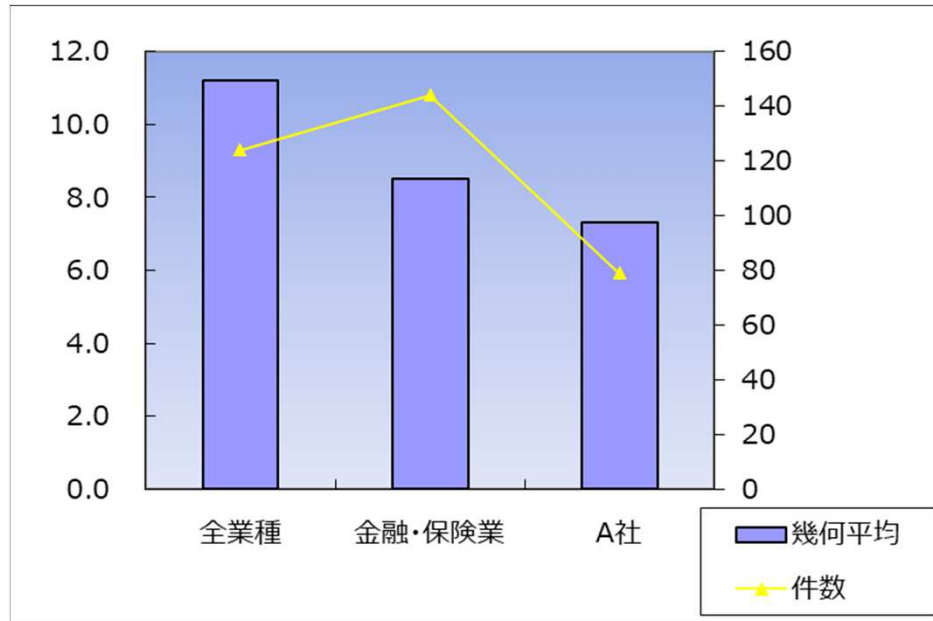
※ 3郡以上の比較では、Tukey-Kramer法などの**多重比較**を適用します。

4.10. 生産性「FP生産性」のPPB比較

31

5工程FP生産性

No	集計区分	件数	幾何平均	標準偏差
1	全業種	124	11.190	2.285
2	金融・保険業	144	8.519	1.974
3	A社	79	7.314	2.083



1. 傾向分析

A社が1人月あたりに開発したFP(幾何平均)は、**7.3FP**です。

- ・ 全業種の**0.65**倍
- ・ 金融・保険業の**0.86**倍

であり、両ベンチマークよりも低い値です。

2. t検定結果(vs 全業種)

P値は**0.02%**で、有意水準5%で有意でした。このため、A社の生産性は「**低い**」と評価しました。

3. t検定結果(vs 金融・保険業)

P値は**12.3%**で、有意水準20%で有意でした。このため、A社の生産性は「**やや低い**」と評価しました。

比較対象	件数	平均	標準偏差	差(A-B)のt検定							
				差(A-B)	自由度	合成分散	合成標準偏差	to値	有意水準	t値	P(両側)
LN(A社)	79	1.98983	0.73378								
LN(IPA全業種))	124	2.41505	0.82648	-0.42522	201	0.626944	0.7917977	-3.73057	5%	1.971837	0.02%

帰無仮説を棄却する

比較対象	件数	平均	標準偏差	差(A-B)のt検定							
				差(A-B)	自由度	合成分散	合成標準偏差	to値	有意水準	t値	P(両側)
LN(A社)	79	1.98983	0.73378								
LN(IPA(金融・保険業))	141	2.1423	0.67995	-0.15247	218	0.489561	0.6996861	-1.55057	20%	1.285447	12.25%

帰無仮説を棄却する

全業種：「グラフデータ 図表6-3-4 ● FP 規模と工数(新規開発)(信頼区間50%, 95% 付き)」を解析して利用

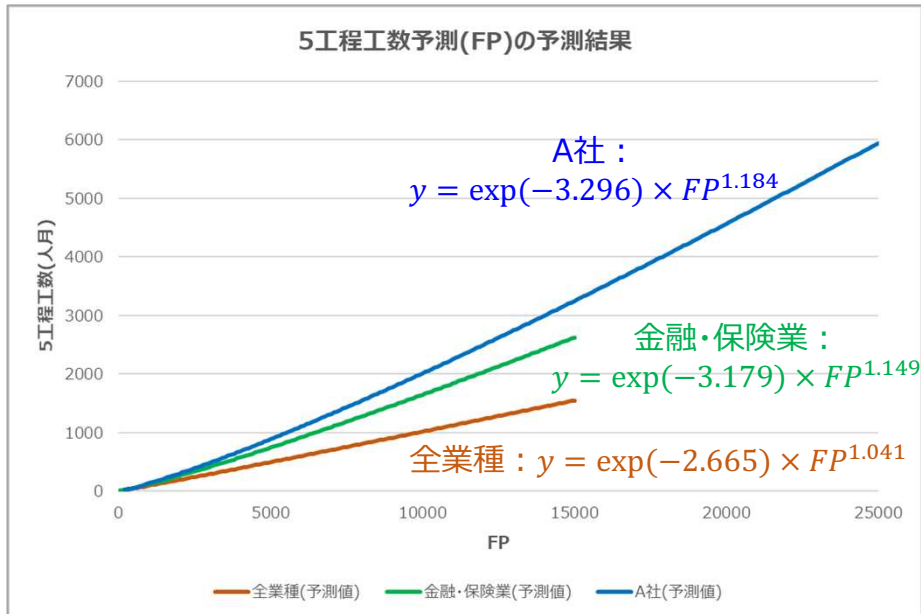
金融・保険業：「グラフデータ 図表6-3-7 ● FP 規模と工数(新規開発, FP 計測手法混在)(信頼区間50% 付き)」を解析して利用

4.11. 生産性「工数予測(FP)」のPPM比較

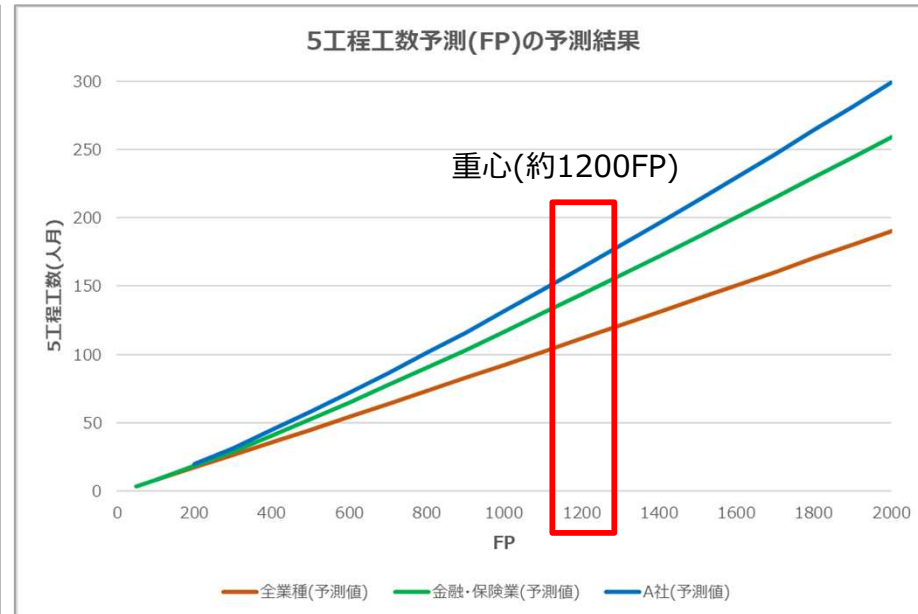
32

5工程工数予測(FP)

No	集計区分	件数
1	全業種	119
2	金融・保険業	134
3	A社	75



全体像



重心付近

低生産性傾向

↑

生産性

↓

高生産性傾向

全業種(オレンジ線)と比較し、A社の開発工数(青線)は、どのFP規模でも低生産性傾向です。実績データの重心(約1200FP)の工数は**1.46倍**であり、A社の生産性は「**低い**」と評価しました。

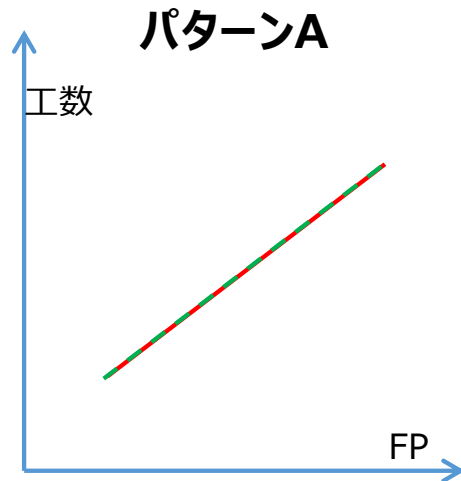
金融・保険業(緑線)と比較しても、どのFP規模でも低生産性傾向です。重心の工数は**1.14倍**であり、A社の生産性は「**やや低い**」と評価しました。

全業種 : 「グラフデータ 図表6-3-4 ● FP 規模と工数(新規開発)(信頼区間50%, 95% 付き)」を解析して利用

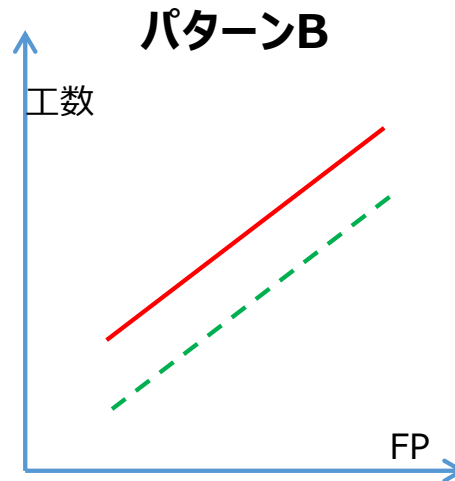
金融・保険業 : 「グラフデータ 図表6-3-7 ● FP 規模と工数(新規開発, FP 計測手法混在)(信頼区間50% 付き)」を解析して利用

【参考】両軸対数散布図での評価

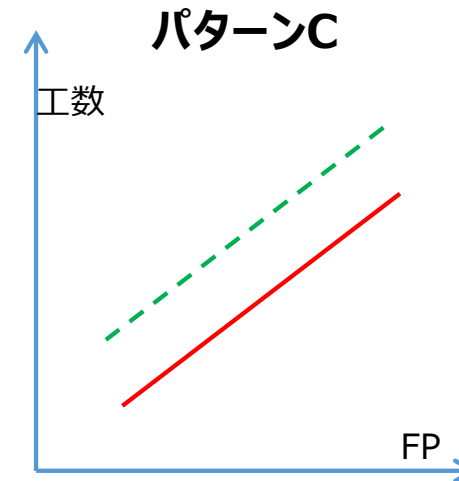
- PPM比較は、両軸対数散布図での確認を推奨します。



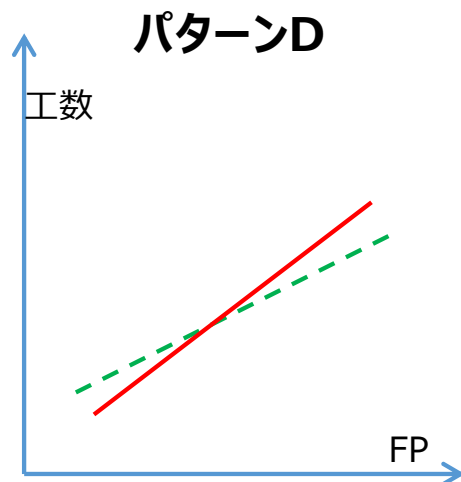
生産性に違いはない。同一の生産性。



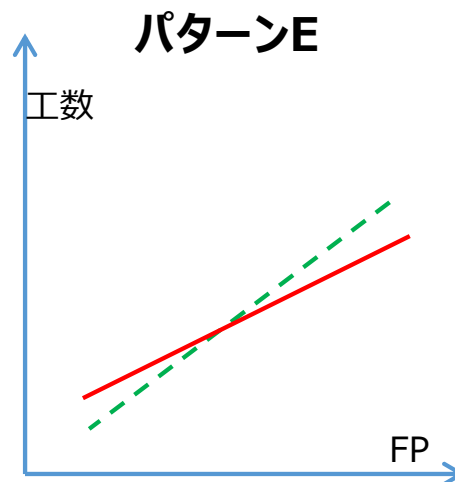
どの規模でもA社の生産性が低い。



どの規模でもA社の生産性が高い。



小規模PJ：A社の生産性が高い。
大規模PJ：A社の生産性が低い。



小規模PJ：A社の生産性が低い。
大規模PJ：A社の生産性が高い。

【凡例】

— A社実績

- - - 外部ベンチマーク

直線で確認するので、直感的に理解することができます。

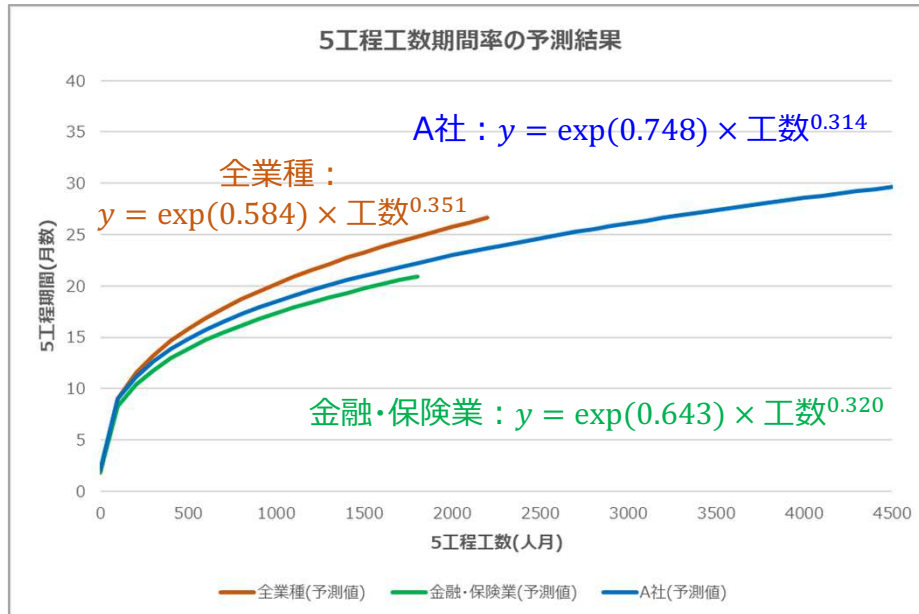


4.12. 工期「工数期間率」のPPM比較

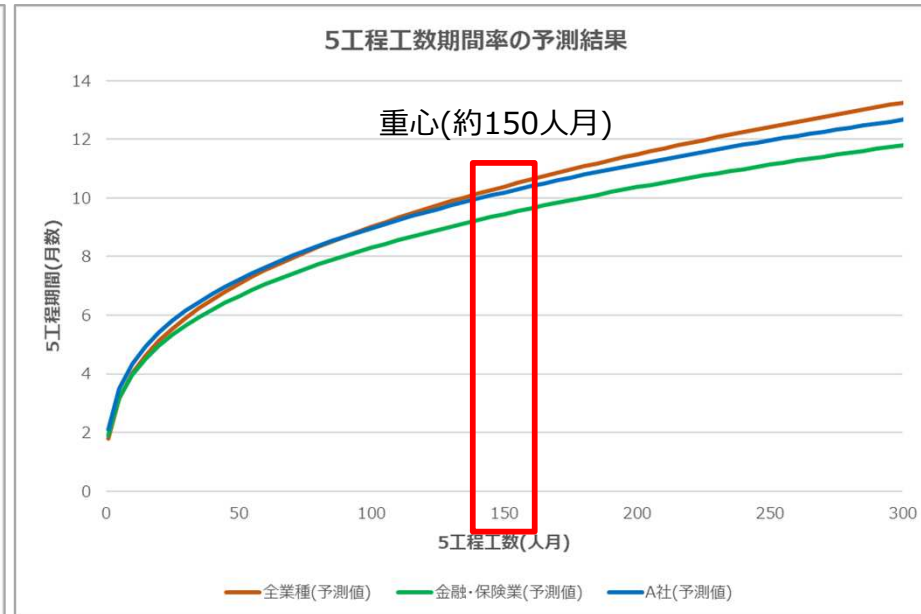
34

5工程工数期間率

No	集計区分	件数
1	全業種	303
2	金融・保険業	218
3	A社	78



全体像



重心付近

長工期傾向
↑
工期
↓
短工期傾向

全業種(オレンジ線)と比較し、A社の開発期間(青線)はほぼ同等です。実績データの重心(約150人月)の工期は0.98倍で、明らかな差は確認できません。

金融・保険業(緑線)と比較すると、わずかですが長工期傾向です。ただし、重心の工期は1.08倍であり、A社の工期は「通常」と評定しました。

全業種：「グラフデータ 図表6-2-2 ● 開発5 工程の工数と工期(新規開発)(信頼区間50%, 95% 付き)」を解析して利用

金融・保険業：「グラフデータ 図表6-2-2 ● 開発5 工程の工数と工期(新規開発)(信頼区間50%, 95% 付き)」を解析して利用

4.13. QCD評価結果のまとめ

- 外部ベンチマーキング結果のまとめは以下のとおりです。

- 考察

- ◆ A社のデータ件数は70件以上であり、傾向は十分に反映できていると想定します。
- ◆ A社のシステム開発は、**品質確保のために、通常より過大な工数を費やしている**可能性があります。

- 品質

- ◆ A社のシステム開発は「全業種」と比較し高品質です。また、「金融・保険業」と比べても高品質の可能性が高いといえます。

- 生産性

- ◆ 「全業種」と比較し低生産性傾向です。また、「金融・保険業」と比較しても生産性が低い可能性があります。
- ◆ 低生産性傾向となった要因については、他メトリクスを用いた追加の調査を実施します。

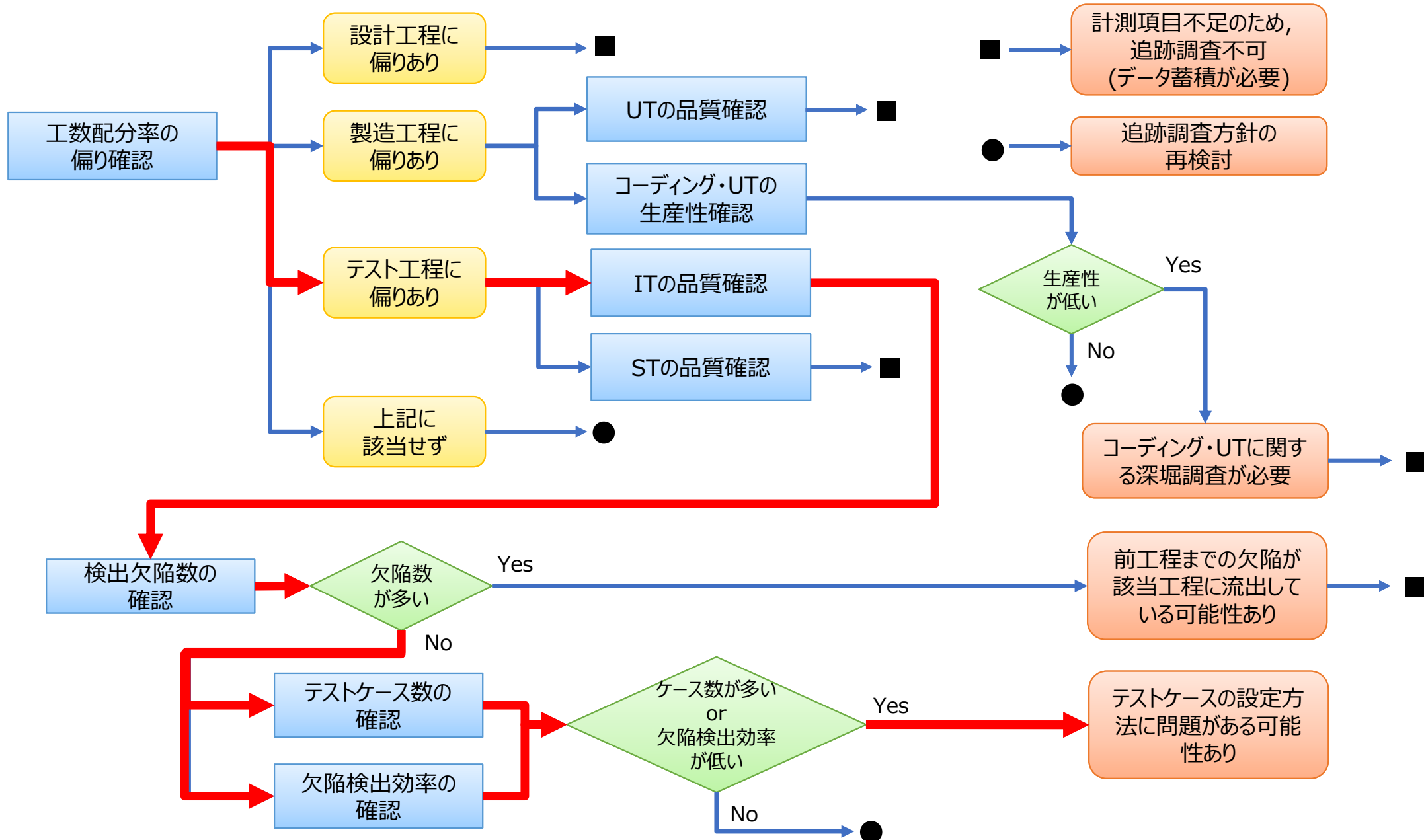
- 工期

- ◆ 開発期間は、「全業種」、「金融・保険業」と同程度で、優劣は特定できませんでした。

5. 外部ベンチマーキングの事例紹介(追跡調査編)

5.1. 低生産性要因に対する追跡調査

- 追跡調査は、「工数配分率の偏り確認」を入り口に実施します。



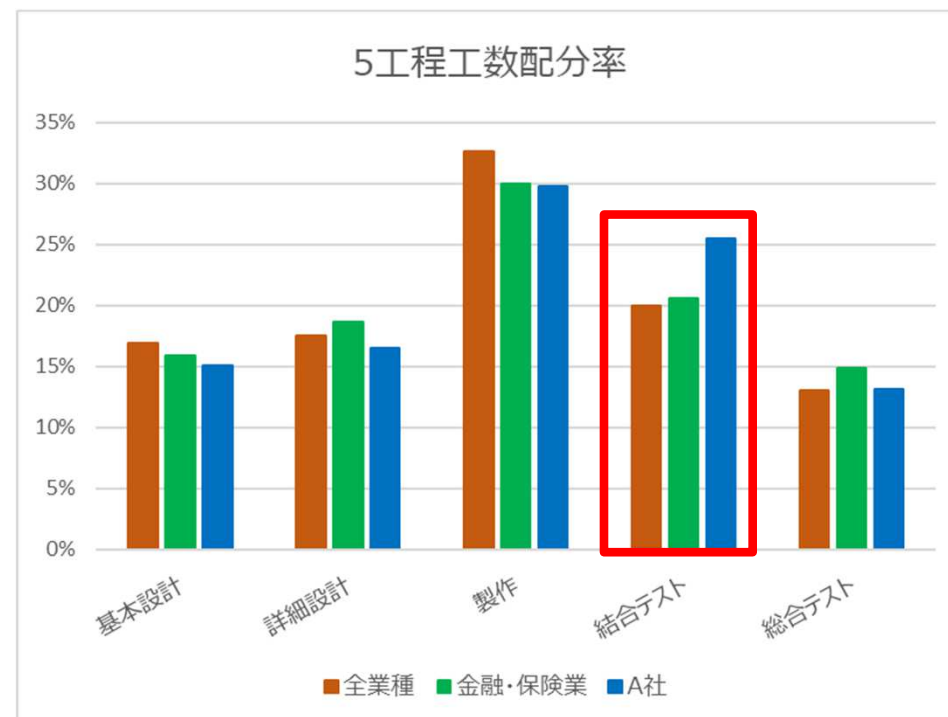
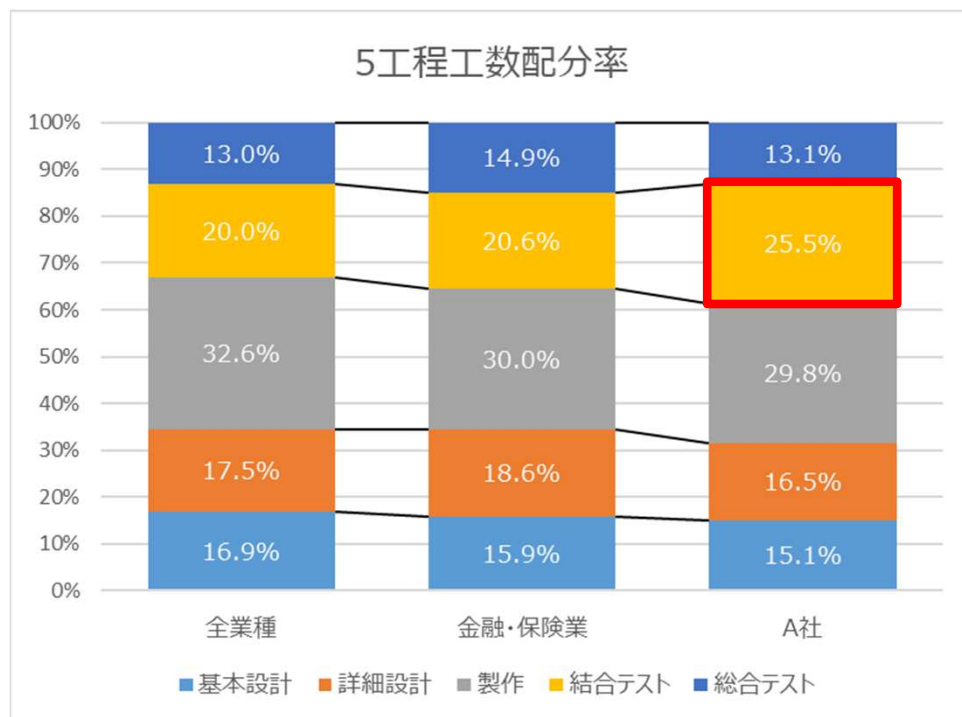
5.2. 追跡調査対象指標

- 追跡調査では，以下の4メトリクスを対象にしました。

No.	確認ポイント	メトリクス(Z)	基本測定量(X, Y)	概要
1	工数配分率の 偏り確認	5工程工数配 分率	Y : (該当)工程工数(人月) X : 5工程工数(人月)	開発5工程の各工程に対する工数配分率 です。 本調査では，配分率に偏りがある工程に何らかの課題があると想定し，確認を進めました。
2	検出欠陥数の 確認	結合テスト欠 陥検出密度 (FP)	Y : 結合テスト検出欠陥数 X : KFP	1,000FPあたりの結合テストの検出欠陥数 です。 結合テストの品質を表す指標として用います。 本調査では，結合テストでの検出欠陥が多く(品質が低い)，改修工数が増大しているのかを確認するために用いました。
3	テストケース数 の確認	結合テスト密 度(FP)	Y : 結合テストケース数 X : FP	1FPあたりの結合テストケース数 です。 結合テストにおけるテスト量が妥当であるのかを確認するために用いました。
4	欠陥検出効率 の確認	結合テスト欠 陥検出効率	Y : 結合テスト検出欠陥数 X : 結合テストケース数 (×1,000)	結合テストケース1,000件あたりの検出欠陥数 です。 結合テストで，欠陥を効率よく検出(ケース自体の品質)できているのかを確認するために用いました。

5.3. 工数配分率の偏り確認

- 外部ベンチマークと比較し、工数配分率に偏りがある工程を確認します。



【分析結果】

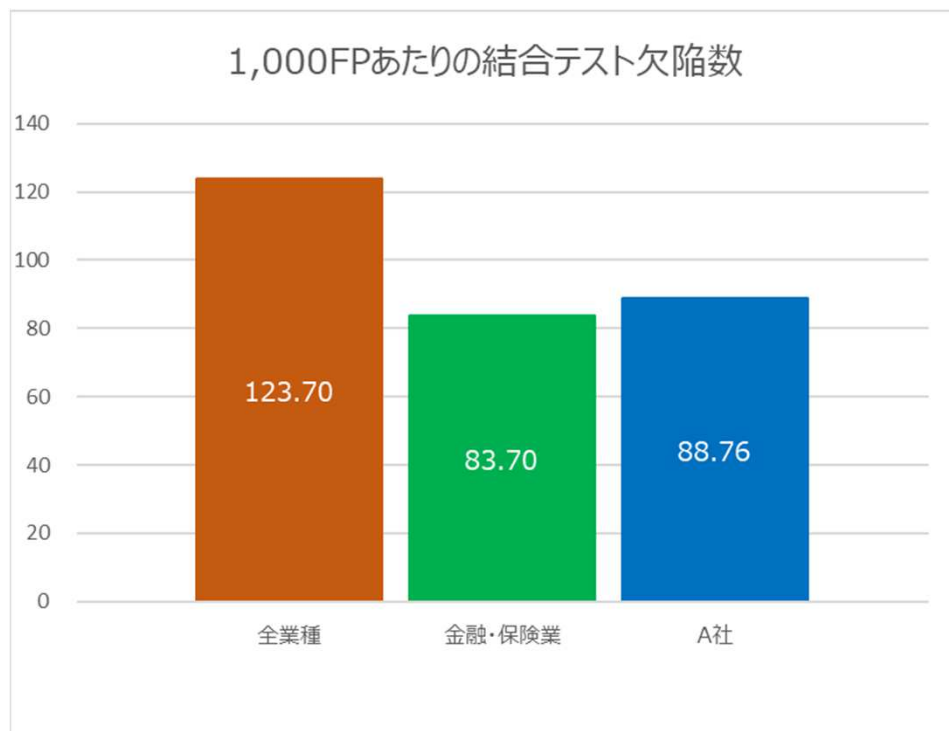
1. 結合テスト工程の工数配分率は25.5%で、全業種と金融・保険業に比べて高い傾向にあります。
2. このため、以降は結合テストの実施状況に着目して調査します。

全業種：「グラフデータ 図表7-1-16 ● 工程別の実績工数の比率の基本統計量(新規開発)」の平均値を利用

金融・保険業：「グラフデータ 図表7-1-13 ● 工程別の実績工数の比率の基本統計量(新規開発)」の平均値を利用

5.4. 結合テスト検出欠陥数の確認

- 前工程までの品質不良で、結合テストの検出欠陥が多い(改修工数が増加している)可能性があるため、欠陥の検出状況を確認します。



【注意】

本メトリクスに対応するグラフデータがないため、データ白書の中央値を利用しています。

このため、t検定による解析はできません。

【分析結果】

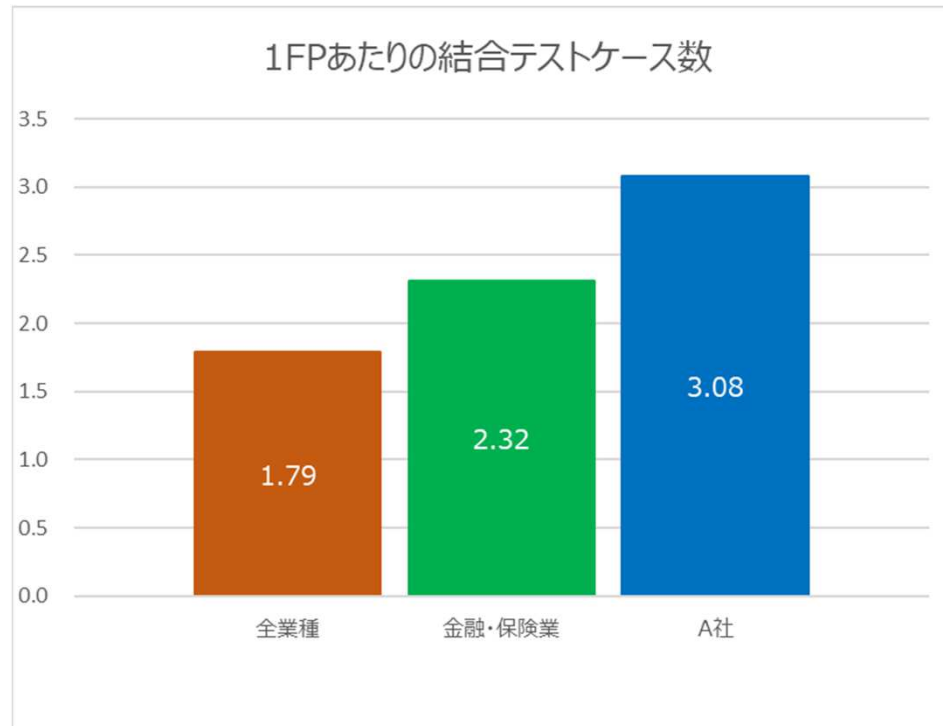
1. 1,000FPあたりの検出欠陥数は、全業種の**0.72倍**ですが、同業の金融・保険業とは同傾向です。
2. 結合テストより前で除去すべき欠陥が過剰に混入した可能性は低く、改修作業に伴う工数の増加はないと想定します。
3. このため、以降は結合テストに費やす労力について調査します。

全業種：「グラフデータ 図表7-5-8 ● テスト工程別FP 規模あたりのテストケース数，検出バグ数の基本統計量(新規開発)」の中央値を利用

金融・保険業：「グラフデータ 図表7-5-8 ● テスト工程別 FP 規模あたりのテストケース数，検出バグ数の基本統計量(新規開発，IFPUGグループ)」の中央値を利用

5.5. 結合テストケース数の確認

- テスト労力が過剰でないかどうかを，結合テストのケース数から確認します。



【注意】

本メトリクスに対応するグラフデータがないため、データ白書の中央値を利用しています。

このため、t検定による解析はできません。

【分析結果】

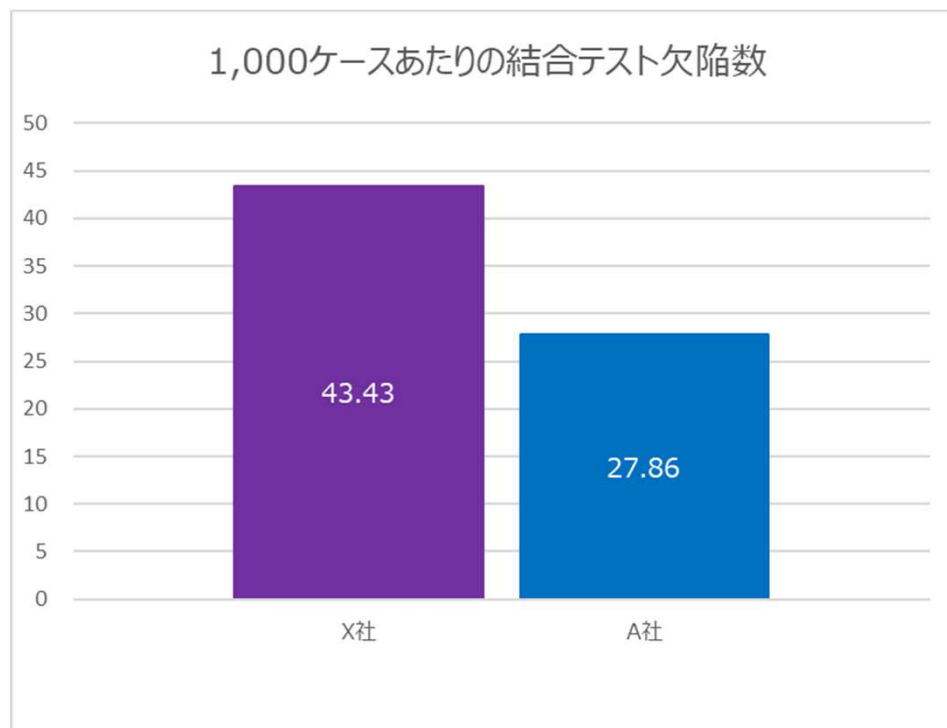
1. 1FPあたりのケース数は，全業種の約**1.72倍**，金融・保険業の約**1.33倍**であり，設定したケース数が多い傾向にあります。
2. 必要以上の過剰なケースの設定など，ケース設定方法が非効率で，低生産性につながっている可能性があります。
3. テストケース設定に関する方針・標準化が不十分の可能性があるため，その記載レベルを確認することを推奨します。
4. 以降は，テストケースに関する欠陥検出の効率性について調査します。

全業種：「グラフデータ 図表7-5-8 ● テスト工程別FP 規模あたりのテストケース数，検出バグ数の基本統計量(新規開発)」の中央値を利用

金融・保険業：「グラフデータ 図表7-5-8 ● テスト工程別 FP 規模あたりのテストケース数，検出バグ数の基本統計量(新規開発，IFPUGグループ)」の中央値を利用

5.6. 結合テストにおける欠陥検出効率の確認

- 結合テストの欠陥検出効率(テストケースの品質)を確認するため、1,000ケースあたりに検出した欠陥数を確認します。



【注意】

データ白書では本メトリクスを扱っていないため、X社のデータを暫定的に利用しています。

【分析結果】

1. 1,000ケースあたりの検出欠陥数は、X社の約**0.47倍**となっています。
2. 上記の結果から、欠陥を効率的に検出するためのケースが設定されていない可能性があります。(単体テストと同等のものや、正常動作の確認が多く実施されている可能性があります。)
3. 各プロジェクトで実施したテストケースの内容確認に加えて、標準化でのテスト工程ごとのケースの設定方針に問題がないか確認することを推奨します。

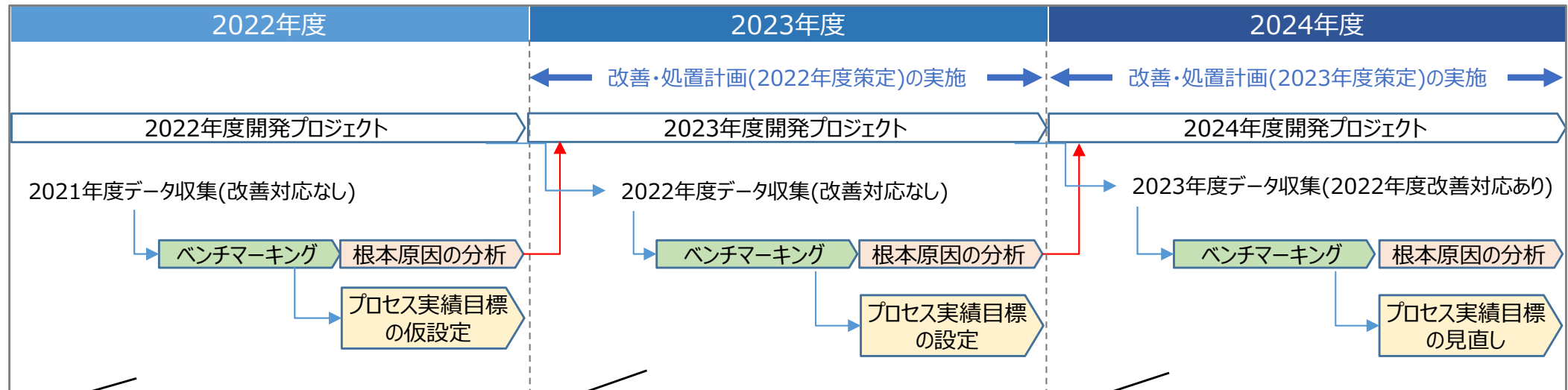
5.7. 追跡調査結果を踏まえた考察

- 今回は、高品質ですが、低生産性傾向と評価しました。低生産性の要因として、結合テストケースの設定方法に改善機会があると推測します。
 - 結合テストのケース数が多いが、ケースあたりの欠陥検出率は低く、非効率なテストの可能性があります。
 - ◆ 通常、テストケースは品質とコストのバランスを保ちながら設定します。
 - ◆ 各システム特性(難易度・複雑度・要求される品質レベルなど)を考慮し、欠陥を効率的に除去するテストケースの設定を推奨します。
 - テストケースに対するレビューで、不要なケースがないことを確認するためのプロセスが必要です。
 - 上流工程で作成する「テスト方針書」の記載レベル・粒度に、問題がないかどうか確認することを推奨します。

【参考】非効率なテストの要因例

- ・ 画面単体での表示やメッセージなど、UTと同等のテストケースを設定している。
- ・ 機能の複雑度・難易度を考慮したカバレッジが曖昧である。
- ・ 入力条件の組み合わせパターンが過剰である。

5.8. 典型的なベンチマーキング結果の活用方法



1. ベンチマーキング

- ・現状実力値, 業界水準値の把握
- ・組織の強み・弱みの特定

2. 根本原因の分析

- ・分析範囲・対象データの選択
- ・問題の根本原因の分析
- ・改善・処置計画(※)の策定
※効果・影響・費用・期間・適用範囲含む

3. プロセス実績目標の仮設定

- ・QCDの相対的優先順位の仮設定
- ・プロセス実績目標の仮設定・周知

改善・処置計画が実施されていない状態では, 現状維持または+α程度の暫定目標の設定に留まることが多い。

1. 改善・処置計画(2022年度策定)の実施

- ・実施結果の評価(定性・定量)
(この段階では定性評価が主体となる。)
- ・中止・変更・適用範囲の拡大検討

2. ベンチマーキング

- ・現状実力値, 業界水準値の把握
- ・組織の強み・弱みの特定
- ・プロセス実績目標の達成度評価(仮)
(改善対象PJのデータが反映できない可能性が高い。)

3. (新たな)根本原因の分析

- ・分析範囲・対象データの選択
- ・問題の根本原因の分析
- ・改善・処置計画(※)の策定

4. プロセス実績目標の設定

- ・QCDの相対的優先順位の設定
- ・プロセス実績目標の設定・周知

1. 改善・処置計画(2023年度策定)の実施

- ・実施結果の評価(定性・定量)
- ・適用範囲の拡大検討

2. ベンチマーキング

- ・現状実力値, 業界水準値の把握
- ・組織の強み・弱みの特定
- ・プロセス実績目標の達成度評価
(この段階で初めて, 改善策の効果が判定できる。)

3. (新たな)根本原因の分析

- ・分析範囲の定義
- ・問題の根本原因の分析
- ・改善・処置計画(※)の策定

4. プロセス実績目標の見直し

- ・QCDの相対的な優先順位の見直し
- ・プロセス実績目標の更新・周知

今どこにいるのか分からなければ、地図は役に立たない。
どこに行きたいのか分からなければ、どの道へ行っても大きな違いはない。

- 外部ベンチマーキングは、以下のために必須です。
 - 企業が進むべき方向を決めるため。
 - プロセス改善効果を把握し、新たなプロセス実績目標を設定するため。
 - (ユーザー企業では)発注の適正化のため。
- SPI担当者(SEPG)のミッションは？
 - 標準プロセスの整備・定着に加え、プロジェクトが要求するメトリクスを提供します。
 - ◆ 統計学では周知の知識や技法を習得し、実績データを正確に解析します。
- プロセス改善機会は、常にプロジェクトからです。
 - プロジェクトに有用なメトリクスの提供が、SPI活動を加速します。

外部ベンチマーキングは、プロセス改善の第一歩です！！

開発プロジェクトの各種見積と定量的マネジメントの実践をご支援

ITメトリクス導入・ 解析支援サービス

IT metrics introduction / analysis support service

ITメトリクス トイックス

検 索

お気軽にお問い合わせください



- 「ソフトウェア開発データ白書2016-2017 金融・保険業編」グラフデータ, 独立行政法人 情報処理推進機構
URL : https://www.ipa.go.jp/ikc/publish/whitepaper_dl.html
- 「ソフトウェア開発データ白書2018-2019」グラフデータ, 独立行政法人 情報処理推進機構
URL : https://www.ipa.go.jp/ikc/publish/whitepaper_dl.html
- 梶山昌之・合田英二・千野智子, 「ソフトウェア開発プロジェクトの計数管理フレームワークによる定量的管理」, プロジェクトマネジメント学会誌Vol.13, 一般社団法人プロジェクトマネジメント学会, 2011年
- 佐藤浩明, 「ソフトウェア開発プロジェクトの生産性評価に関する事例 ―真の価値と工数削減効果を計測する手法」, ユニシス技報 2016年9月発刊 Vol.36 No.2 通巻129号, 2016年
URL : https://pr.biprogy.com/tec_info/tr129/129abs.htm
- 佐藤浩明, 「SPI活動と連携した定量的プロジェクトマネジメントの実践」, SEC journal No.51 第13巻 第3号(通巻54号), 2017年
URL : <https://www.ipa.go.jp/sec/secjournal/051.html>
- 佐藤浩明, 「ソフトウェア開発プロジェクトにおける工数見積りの妥当性評価手法」, ユニシス研究会, 2017年