

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

Case Study for the Productivity Evaluation of Software Development Projects —— Technique to Measure the True Value and Man-hour Reduction

佐藤 浩明

要約 生産性向上に伴う開発コストの削減は、多くのIT企業にとって共通の目標である。したがって、生産性評価の主目的は、生産性向上による開発工数の削減量を確認することでもある。しかし、IT分野の代表的な生産性指標であるFP生産性では、プロジェクトの真の価値を完全に表現できず、生産性の変化に伴う工数の削減効果が計測できない。この問題を解決する新たな生産性指標として「標準工数」を用いた「工数生産性」を採用した。本稿では、FP生産性の問題点と、工数生産性を評価するためのツールである「工数生産性評価モデル」の概要について紹介する。

Abstract Software development cost reduction based on productivity improvement is a common goal for many IT companies. Therefore, the main purpose of productivity evaluation is to confirm the reduced amount of development man-hour result from the productivity improvement. However, FP productivity, known as the typical productivity indicator in the area of the IT, can't fully reflect the true value of the project, so we can't measure the effect of reduction effort. In order to solve this problem, we proposed to use a productivity index, 'Man-hour productivity', based on standard man-hour. In this paper, some problems related to FP productivity are described, and we introduce 'Man-hour productivity evaluation model' which is a tool for evaluating the man-hour productivity.

1. はじめに

東北インフォメーション・システムズ株式会社（TOiNX：以降、当社）は、2001年からCMMI[®]（能力成熟度モデル統合^{*1}）を用いたソフトウェア開発プロセスの改善活動に取り組んでいる。この改善活動により、システム稼働後の流出欠陥数が減少し、品質の向上を定量的に確認している。さらに昨今は、現行の品質水準を維持しつつ、生産性を向上するための施策に取り組んでいる。この施策の有効性を判断するため、開発組織全体と各プロジェクトの生産性をFP（ファンクション・ポイント）を用いたFP生産性で分析した。しかし、この過程でIT分野の代表的な生産性指標であるFP生産性には、重大な問題が内在しており、生産性を正確に評価できないという事実が判明した。この指標では、開発プロジェクトの真の価値が表現できず、生産性の変化に伴う工数の削減効果も計測できないのである。

この問題を解決するため、「標準工数」を用いた生産性指標である「工数生産性」を採用し、生産性の変化と工数削減効果の両方を計測した。本稿では、2章で生産性と指標の定義、3章でFP生産性の問題点を説明し、4章から7章までで工数生産性評価モデルの概要、構築方法、評価、今後の方向性について述べる。

2. 生産性指標の定義

まず、生産性の本来の定義を確認すると共に、IT分野の代表的な生産性指標が持つ特徴について記述する。

2.1 生産性とは

生産性とは、あらゆる分野で共通に利用する用語である。一方で多くの曖昧さを含んだ用語でもあり、対象分野によって多様な算出方法や意味合いで用いる。だが、その本質的な定義は明確で「成果量」と「投入量」の関係を表す指標である。

$$\text{生産性} = \frac{\text{「成果量」 (規模尺度：個数, 重量, 面積, 複雑さ など)}}{\text{「投入量」 (費やした工数, 時間, 人数, コスト など)}}$$

一般的に、同一の製品を繰り返し製造する分野では、「1時間あたりに100個の製品を製造できる」といったように生産性を表現する。もし、何らかのプロセス改善効果で、この能力が200個になった場合、生産性は2倍に向上したと評価でき、同時に、経営に対するプラスの効果を期待する。生産性の向上に伴い、製造に費やすコストや時間が削減できるためである。このように本来の生産性とは、企業経営に対する影響や効果に直結する指標である。しかし、次節に記すFP生産性は、IT企業の経営に対する影響や効果を説明できない。

2.2 IT分野における生産性指標

独自の成果物を創造するソフトウェア開発プロジェクトでは、成果量を示す規模尺度が少ない。このため、FP生産性と開発ライン数のSLOC (Source Lines Of Code) を用いたSLOC生産性の二つが、IT分野の代表的な生産性指標となる。FPには、IFPUG法^{*2}という厳密な計測ルールが存在する。しかし、計測作業自体の負荷が高く、FPをカウントしたプロジェクトが少ない傾向にある。一方、後者のSLOCには規模尺度として幾つかの問題があり、特に以下の一つ目が、生産性評価の観点で致命的である。

- 1) エンハンス (機能改良) は、既存プログラムの数ラインの修正に対して、広範囲な母体システムをテストする場合があります、SLOCと工数の関係が弱い。
- 2) 開発言語別の厳密なSLOC計測ルールが業界で統一されていない。
- 3) 単一言語による開発は少なく、複数言語のSLOCをどのように扱うのが難しい。

FPとSLOC生産性は、「ソフトウェア開発データ白書^{*3}」や、「ソフトウェア開発データリポジトリの分析^{*4}」などの外部情報が存在する。このため、自社内の生産性評価のみならず、業界水準値との比較が可能である。しかし、外部情報のSLOC生産性には、当社のメインフレーム系の開発で利用するPL/I言語が対象外であるなど、言語の多様性に対応できないという問題もある。

なお、国内の外部情報では、「共通フレーム (SLCP-JCF)^{*5}」で定義された、基本設計から総合テスト (ベンダ確認) 工程までの、開発5工程の実績工数を用いて生産性を算出する。これは、要件定義と総合テスト (ユーザー確認) 工程の工数は、対象業種によってバラツキが大きいとするためである。しかし当社では、要件定義からシステムの移行を含む、すべての開発

工程が作業の対象となる。このため、自社向けの生産性評価には全開発工程の工数を、外部情報との比較には開発5工程の工数を利用するといった使い分けが必要になる。

3. FP 生産性の分析結果と特定した問題

当社は、SLOC が持つ規模尺度としての問題を考慮し、FP 生産性を採用した。ここでは、FP 生産性の分析方法、分析結果、及び特定した三つの重大な問題を記述する。

3.1 FP 生産性の分析方法

FP 生産性とは、開発完了時の FP 値を実績工数で除算したもので、単位工数あたりに製造した FP を表す。算出式は単純で各プロジェクトの生産性の計測は容易である。だが、開発組織全体を計測する場合は、やや複雑な手順になる。

これは、FP と実績工数の双方が「対数正規分布」という確率分布に従うためである。さらに、対数正規分布どうしを除算した結果も対数正規分布になるため、平均値の扱いに注意が必要となる。図1は、当社の FP 生産性のヒストグラム^{*6}である。横軸は、生産性を一定の間隔で区分した階級で、右側に行くほど高生産の階級になる。縦軸は、その階級に該当するプロジェクト数を示す。対数正規分布は左右非対称で、右裾に緩やかに広がった形状になる。このため、各プロジェクトの FP 生産性の総和をプロジェクト数で除算した、通常の意味の平均値（算術平均値）が利用できない。図1の矢印のように、算術平均値は分布の右裾に位置する少数の高生産性プロジェクトの影響を受け、プロジェクト数が多い階級から右側にずれる。このため、算術平均値では、組織全体という集団の特性を正確に表現することができない。

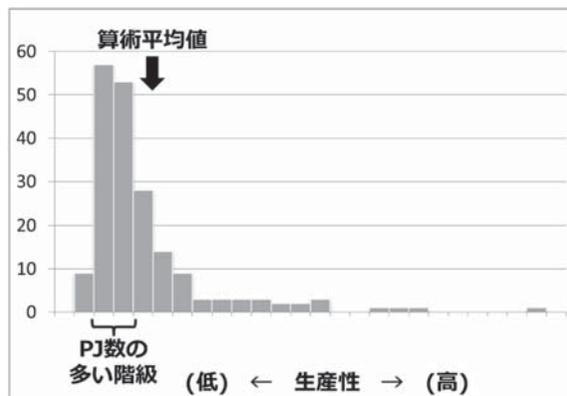


図1 FP 生産性分布

対数正規分布には、データを対数変換^{*7}することで「正規分布」に近似するという性質がある。図2は、図1と同じデータを対数に変換した後のFP 生産性のヒストグラムである。分布の女王と呼ばれる正規分布は、医学や社会科学などの多くの分野で利用され、その性質が昔から研究されている。分布は左右対称のベル型を描き、「算術平均値 = 中央値^{*8} = 最頻値^{*9}」になるため、算術平均値が分布の代表特性値となる。また、データのバラツキを示す統計量の「標準偏差 (σ)」を用いると、算術平均値を中心とした左右の一定区間に含むデータの確率が計算できる。

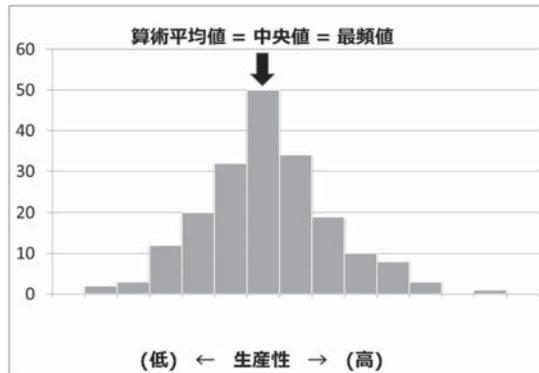


図2 対数変換後のFP生産性分布

分布が正規分布であるかどうかは、各種の正規性の検定手法で確認できる。ただし、実用的な分析では、Excel関数で準備されている分布の対称性を示す「歪度（わいど）」と、鋭さを示す「尖度（せんど）」という統計量の値が ± 1.5 未満かどうかで判定する方法で十分である。

当社のメトリクス^{*10}分析では、SLOC、テストケース数、欠陥数など、多くのITデータが対数正規分布に近似することを確認している。FP生産性の分析においても、対数変換後の正規分布の性質を利用することで、適切な基準値と上下限値が設定できる。

ただし、対数変換後に算出した値は元の値と単位が異なるため、この数値自体に意味はない。そこで、図3のように、正規分布で求めた算術平均値と ± 1 標準偏差（ σ ）を指数関数^{*11}で元の単位に戻す。この平均値は「幾何平均値」と呼ばれ、組織全体の特性を表す基準値となる。また、 ± 1 標準偏差の範囲には、全データの約2/3を含む性質があり、生産性の正常範囲を判定するための最適な上下限値となる。正規分布では算術平均値から左右に等距離であった標準偏差は、対数から元の単位に戻す際にマイナス側の幅は狭く、プラスの幅は広くなる。

さらに統計学には、AとBという二つの集団の平均値に、意味のある変化があったのかどうかを判定する「t検定」という手法がある。生産性向上施策の適用前後の組織全体の平均値を検定することで、生産性に意味のある変化があったのかどうかを確認できる。

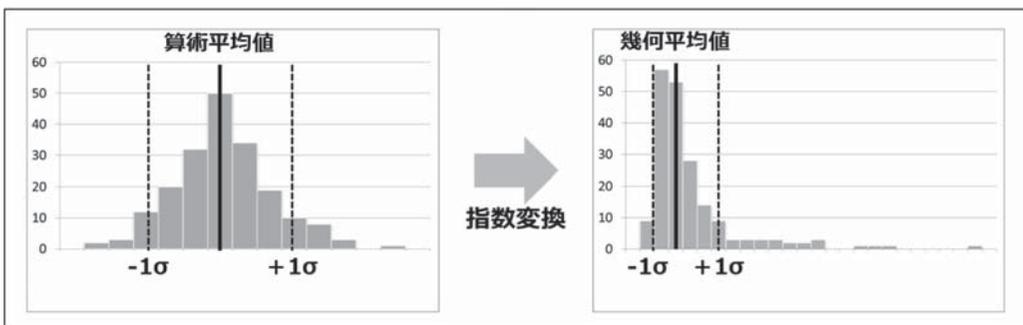


図3 指数関数による基準値の設定

3.2 FP 生産性の分析結果

蓄積した 250 の実績データから、FP 値（当社オリジナルの FP 推測値も併用）と、実績工数の両値を持つ約 200 プロジェクトを対象に分析した結果、生産性の基準値は、概ね 10FP/人月であった。しかし、正常範囲の目安である ± 1 標準偏差は、約 4FP/人月から 26FP/人月であり、6 倍以上の幅を特定した。また、個々のプロジェクトの値を確認すると、最小は 1FP/人月未満で、最大は 100FP/人月を超えていた。基準値を中心に 1/10 倍から 10 倍以上という、実に 100 倍以上の開きがあった。

このバラツキの大きさから、t 検定では施策の適用前後の生産性に、意味のある変化を確認することができなかった。当然、施策の適用前後の FP 生産性は異なる値である。しかし、統計学に基づく結論では、二つの生産性の違いは誤差の範囲であり、生産性が向上、又は下降したと断定できないのである。単純に、生産性の向上施策を適用した各プロジェクトの FP 生産性と基準値を比較することも試みた。しかし、顧客を含む関係者から高評価を得た成功プロジェクトが、基準値の半分にも満たないケースや、やや苦戦気味であったはずのプロジェクトが、基準値の数倍の高生産になるといった事象が発生した。

分析の対象を新規開発やエンハンスなどの開発タイプで層別^{*12}すると、バラツキは多少であるが狭まる。併せて、主 OS や主開発言語、開発規模別で層別を繰り返すと、バラツキは、より小さくなる。しかしこの場合、対象データ数が不足し、平均値自体の信頼性が低下する。エンハンスの Java など、データ数が十分にある層別も分析したが、依然としてバラツキは大きく、t 検定で有意差は確認できなかった。

さらに、このバラツキの大きさに関する問題は、SLOC 生産性も同様であった。つまり、IT 分野の代表的な規模尺度を用いた FP や SLOC 生産性では、組織全体と各プロジェクトの生産性を評価することができないのである。

3.3 FP 生産性の三つの問題

FP 生産性には「成果量の網羅性の低さ」、 「プロジェクト規模特性の未反映」、及び「プロジェクト構成の影響を受ける」という根本的な問題がある。この三つの問題について以降に記述する。

3.3.1 成果量の網羅性の低さ

FP は、対象ソフトウェアの機能規模を表す尺度である。しかし、この尺度だけでは、プロジェクトの真の価値である成果量のすべてが網羅できない。図 4 に示すように、FP は成果量の一部を説明する。しかし、成果量には要求される品質レベル、開発言語などの技術特性、及びお客さまの特性なども反映しなければならない。さらに、成果量に占める FP の割合はプロジェクト特性によって異なり、FP だけでは一部の成果量しか説明できない場合と、大部分の成果量を説明できる場合がある。

開発プロジェクトの真の価値とは、複雑に絡み合う複数の要因を反映した成果量のことであり、FP や SLOC などの単一の規模尺度で単純に表現できないのである。

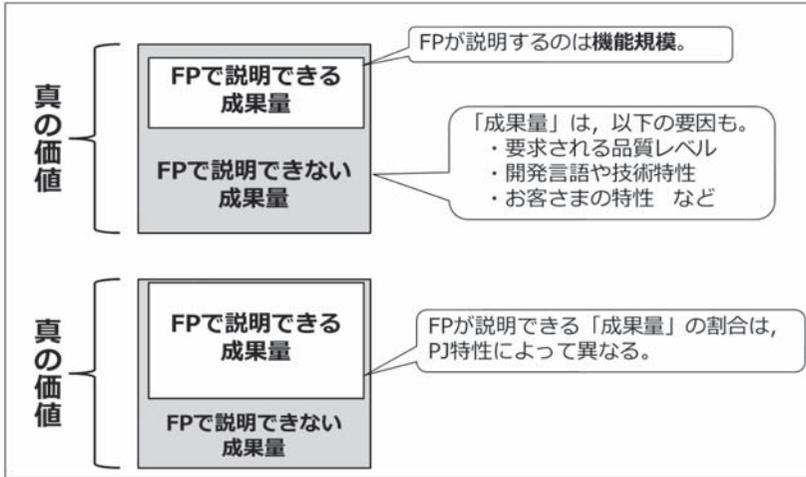


図4 FPと成果量の関係

例えば、同じ1,000FPのAとBのプロジェクトが存在したとする。Aの実績工数は100人月で10FP/人月、Bは125人月で8FP/人月の場合、単純にBが低生産性と評価される。ところがBプロジェクトは、多数のユーザーを対象とする入金や金額請求を扱うシステムの開発であった。極めて高いセキュリティと品質レベルが要求され、Aプロジェクトの10倍以上のテストケースを設定して品質を確保していた。しかし、機能規模だけを表すFPでは、この要求に対応する労力を成果量として反映できないのである。同様に、非常に複雑な内部ロジックが多いが、その内容が画面や帳票などの入出力機能に反映されないシステムの場合、平均的な業務システムの1/10程度のFP値になることがある。当然、FP生産性は低くなるが、開発プロジェクトの能力に問題がある訳ではない。

3.3.2 プロジェクト規模特性の未反映

FP生産性は、大規模や小規模といった開発プロジェクトの規模特性が反映できない。図5の左側に示すように、一般的な知見として、FPの増加に伴い開発工数は徐々に増加傾向を強め、曲線的に増える。これは当社のメトリクス分析でも同様の傾向を確認している。

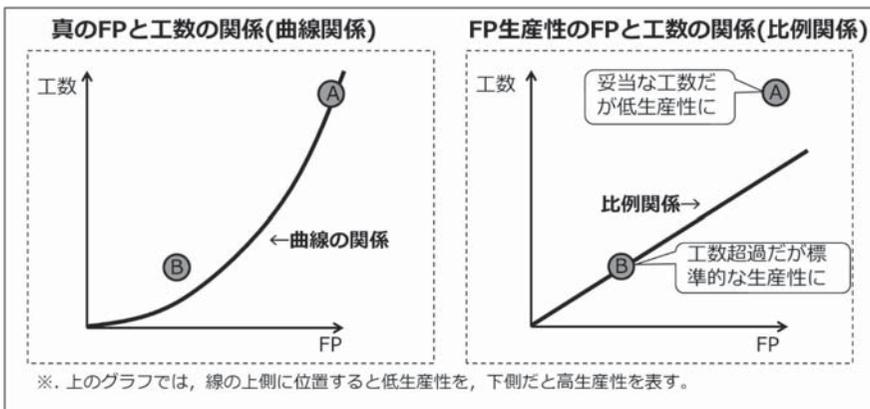


図5 規模と工数の関係

ところが、FP 生産性は図 5 の右側のように、FP 値と実績工数を直線の比例関係で考えるため、各プロジェクトの規模に応じた生産性評価ができない。このため、規模的には妥当な工数を費やした大規模プロジェクト A を低生産性と評価したり、低生産性な小規模プロジェクト B を標準的な生産性と判断したりする危険がある。例えば、10FP/人月が基準値の場合、10,000FP では 1,000 人月が妥当となる。しかし現実的には、プロジェクトの規模が大きくなるにつれて複雑度やインターフェースが飛躍的に増加し、より多くの工数を費やすことが明白である。

3.3.3 プロジェクト構成の影響を受ける

FP 生産性で組織全体を評価する場合、対象年度や組織のプロジェクト構成によって、異常な結果になることがある。各プロジェクトの生産性の変化と、組織全体の生産性の変化が矛盾するという事象である。なお、組織全体の FP 生産性は幾何平均値が基準値となるが、算出方法がやや複雑になるため、以下の例では算術平均値を用いて説明する。

図 6 の上部は、Web 系の開発プロジェクトを高生産性タイプ、メインフレーム系の開発プロジェクトを低生産性タイプとした例である。なお、各タイプの生産性は変化せず、高生産性タイプが 15FP/人月、低生産性タイプは 5FP/人月で一定とする。N 年度は、高生産性タイプが 3 件、低生産性が 2 件で、組織全体の生産性は 11FP/人月である。ところが、N+1 年度のように、この割合に変化が生じると、各タイプの生産性に全く変化がなくても、全体の生産性は 9FP/人月に下降する。

また、図 6 の下部のように、何らかのプロセス改善効果で、すべてのプロジェクトの生産性が向上した場合でも、組織全体では低下するというパラドックスが起こる。例えば、ある改善効果で、高生産性タイプが 15FP/人月から 16FP/人月に、低生産性タイプが 5FP/人月から 6FP/人月に一律に向上したとしても、組織全体の生産性は 10FP/人月に低下する。

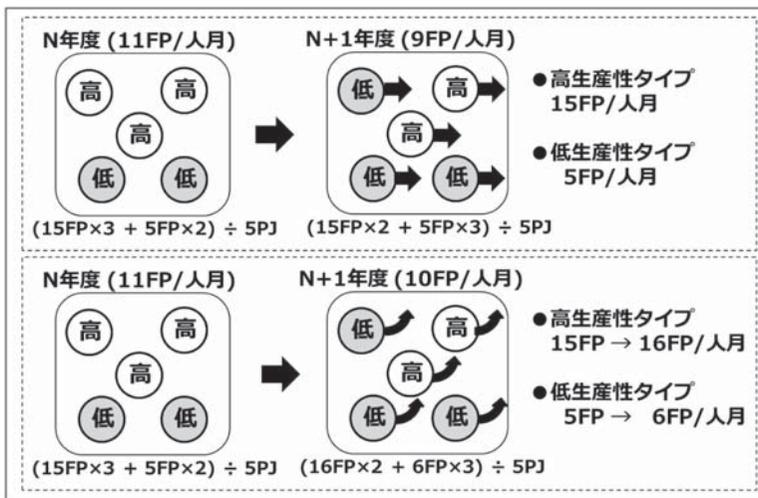


図 6 プロジェクト構成による影響

この事象は、生産性低下の場合だけではない。真の生産性向上を果たしていないにもかかわらず、生産性が向上したと判断するケースもある。このため、生産性の向上施策に対する有効

性を客観的に判断することができない。

このように、FP生産性には幾つかの根本的な問題が内在している。この指標を開発プロジェクトの生産性評価に用いると妥当な判断ができない。これはFP生産性が、経営者やプロジェクトの関係者が期待する、企業経営に直結した本来の生産性指標の定義と異なるためである。しかし、現在のIT分野では、FP生産性が持つこの問題が十分に議論されていない。

4. 新たな生産性評価指標の導入

FP生産性の問題を解決する新たな生産性指標である「工数生産性」の概念と、この生産性を評価するためのツールである「工数生産性評価モデル」の特徴について記述する。

4.1 工数生産性の概念

生産性の評価には、プロジェクトの真の価値を示す成果量が必須である。しかし、この成果量を、FPやSLOCという単一の規模尺度に限定する必要はない。各プロジェクトの特性に応じた基準となる真の価値を算出し、成果量とすればよいのである。このとき、蓄積した実績データから、対象プロジェクトの特性に応じた基準となる「あるべき工数」が予測できれば、この値は真の価値を示す「標準工数」となる。そして、この標準工数は、生産性指標の成果量として利用できる。この生産性指標を「工数生産性」と定義した。

標準工数と実績工数を比較すると、各プロジェクトや組織全体の生産性の変化だけではなく、工数の削減効果も計測できる。評価結果は、以下の単純な式で算出でき、工数生産性は値が1より大きい場合に、工数削減効果は0より大きい場合にプラスの効果があったと判定する。

【各プロジェクト (PJ) の評価】

$$\text{工数生産性} = \text{標準工数} \div \text{実績工数}$$

$$\text{工数削減効果} = \text{標準工数} - \text{実績工数}$$

【開発組織全体の評価】

$$\text{工数生産性} = \text{各PJの標準工数の総和} \div \text{各PJの実績工数の総和}$$

$$\text{工数削減効果} = \text{各PJの標準工数の総和} - \text{各PJの実績工数の総和}$$

$$\text{平均工数生産性} = \text{各PJの工数生産性の総和} \div \text{PJ数}$$

例えば、あるプロジェクトの標準工数が100人月で、実績工数が80人月の場合、工数生産性は「1.25」となる。これは、過去のプロジェクトと比較して1.25倍の効率で作業したことを意味し、生産性が25%向上したことになる。また同時に、20人月の工数を削減したと評価できる。同様に、ある年度の組織全体を評価したとする。このとき、該当年度に完了したプロジェクトの標準工数の総和が1,000人月、実績工数の総和が900人月とすると、工数生産性は「1.11」になる。これは、基準となる年度と比較して11%の生産性向上を果たしたことを意味し、同時に、組織全体で100人月の工数削減を達成したと評価する。

ただし、組織全体の工数生産性は、大規模なプロジェクトが極端な高生産性（又は低生産性）になると、その影響を受けて、小規模プロジェクトの結果を打ち消す。このため、生産性向上施策の有効性といったプロセス改善効果を判定する際には、プロジェクト規模の大小による重みを反映しない平均工数生産性が妥当である。

このように工数生産性は、基準とする過去の実績値と比較し、評価時点での実力値の変化と、工数の削減量を如実に表すため、ビジネス上の効果に直結した生産性指標となる。

4.2 工数生産性評価モデルの特徴

標準工数を予測し、工数生産性を評価する工数生産性評価モデルとは、多変量解析^{*13}の一種である重回帰分析を活用したツールである。重回帰分析とは、あるデータ（目的変数^{*14}）を、他の複数のデータ（説明変数^{*15}）で予測する関係性の式（回帰式）を作る手法である。適切な説明変数を選択することで、説明変数が一つだけの単回帰分析よりも高精度な予測ができる。そして、この評価モデルが持つ特徴は、3.3節で述べたFP生産性の三つの問題を解決する。

4.2.1 成果量の網羅性が高い

この評価モデルは、各プロジェクトの特性に応じて、表1の区分と説明変数（規模尺度）から、図7に示すとおり標準工数を最も説明できる層別の回帰式を自動で選択する。

表1 区分と説明変数（規模尺度）

●区分	
区分名	解説
開発タイプ	新規開発, 再構築, コンバージョン, エンハンス, メンテナンス など
主言語	Java, PL/I, VB系, C系 など
主OS	Solaris, Windows, AIX, Z/OS など
顧客	顧客の企業名や部門名
●説明変数（規模尺度）	
説明変数名	解説
FP	FP値, または, 画面数や帳票数などから算出した当社オリジナルのFP推測値
SLOC	空白行, コメント行を除く物理行数
UT	単体テストで実施した総テストケース数
IT	サブシステム・リンクテストと統合テストで実施したテストケース数の合計値
ST	システムテストで実施した総テストケース数

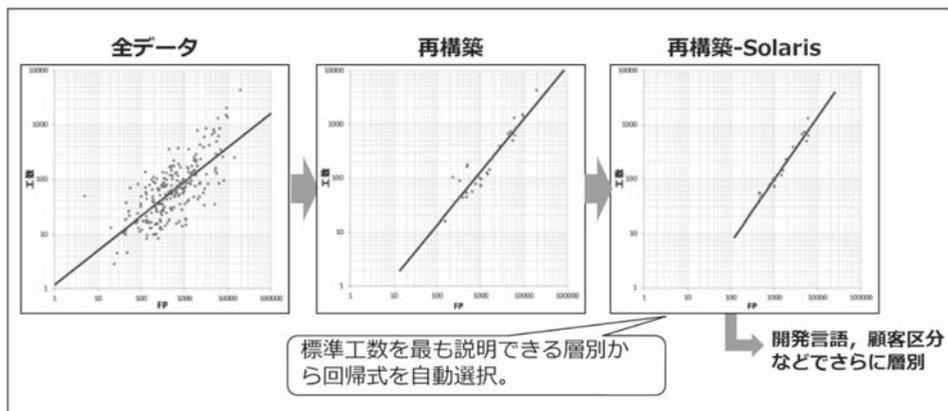


図7 回帰式の選択

また、説明変数は、FP と SLOC に加えて各テストケース数も利用しており、品質要求を反映した標準工数を算出する。例えば、図 8 のように「新規開発-Java-Solaris-顧客 A」を指定すると、同じ特性を持つ過去の実績データから作成した「FP、統合テストケース数、システムテストケース数」の三つの説明変数を用いた回帰式を自動で選択して標準工数を導き出す。同様に、「メンテナンス-PL/I-Z/OS-顧客 B」を指定すると、「SLOC、単体テストケース数、統合テストケース数」で標準工数を予測する。このため、FP や SLOC だけでは説明できない、複雑な要因が絡み合うプロジェクトの真の価値を、標準工数という単一の尺度で網羅できる。



図 8 プロジェクト特性に応じた成果量の網羅

4.2.2 プロジェクト規模特性の反映

非常に大規模なプロジェクトの場合、FP 生産性は基準値の数分の 1 から数十分の 1 になることがある。これは FP 生産性が、プロジェクトの規模特性を無視した指標だからである。しかし本評価モデルは、図 9 の破線で示す「べき乗回帰」という曲線を描く回帰式を搭載している。このため、FP 生産性のような直線の比例関係では扱うことができない、大規模プロジェクトの生産性の低下傾向を反映した標準工数が算出できる。

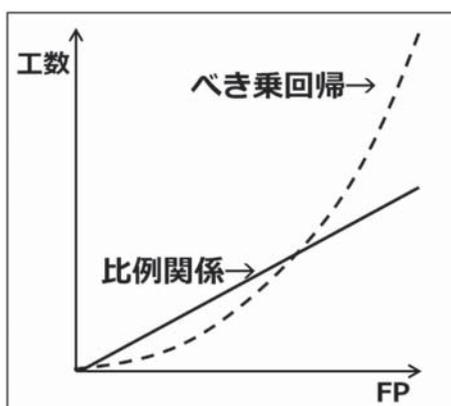


図 9 べき乗回帰による規模特性の反映

4.2.3 プロジェクト構成の影響を受けない

この評価モデルで算出する標準工数は、対象プロジェクトの技術特性、品質要求、規模特性、顧客特性を反映するため、組織や対象年度のプロジェクト構成に影響を受けない。例えば、あ

る年度に、低生産性になりがちなメインフレーム系の大規模プロジェクトが集中しても支障はない。標準工数は、過去の同様の特性を持つプロジェクトの実績データに基づいて算出するため、プロジェクト構成の偏りに起因する誤分析は発生しない。つまり、どのようなプロジェクト構成であっても、生産性の変化と工数削減量の両方を計測できるのである。

5. 工数生産性評価モデルの構築方法

本章では、工数生産性評価モデルを構築するための、三つの主要なデータ解析手順を記述する。

5.1 層別の検討

実績データの件数を考慮しながら、開発タイプや主言語などの要因を組み合わせることで、代表的な規模尺度であるFPとSLOCで予測できる工数の割合を限界まで向上させる層別を探し出す。この作業は、どの層別が最適であるのかが事前に想定できないため、要因の統合と分離を何度も繰り返して候補を選定する。今回の解析作業では、データ数と工数の予測精度の観点から89の層別候補を設定した。

5.2 重回帰分析の実施

設定した層別ごとに、FP、SLOC、及び単体テスト、統合テスト、システムテストのテストケース数を説明変数として組み合わせた重回帰分析を実施する。成果量の一部を説明するFPやSLOCに、品質確保の労力を表すテストケース数を加えることで、成果量の網羅性を高めるためである。今回は、89の層別候補に対して表2に示す15タイプの計1335パターンの重回帰分析を実施した。なお、この15タイプは、生産性評価の対象プロジェクトにおける、一部データの欠落、又は未計測を想定して設定したものである。つまり、FPやSLOCなどの一部データが欠落している場合であっても、生産性を評価可能にするためのものである。

表2 重回帰分析のタイプ

No.	タイプ	説明変数 (括弧内は短縮名)				
		FP(F)	SLOC(L)	UT数(U)	IT数(I)	ST数(S)
1	FLUIS	○	○	○	○	○
2	FLIS	○	○		○	○
3	FL	○	○			
4	FUIS	○		○	○	○
5	FIS	○			○	○
6	F	○				
7	LUIS		○	○	○	○
8	LIS		○		○	○
9	L		○			
10	UIS			○	○	○
11	UI			○	○	
12	IS				○	○
13	U			○		
14	I				○	
15	S					○

○：該当するタイプで説明変数として取り込むもの

通常、重回帰分析には、複雑な操作を要する専用のソフトウェアを用いる。しかし、当社が2012年に開発した「マトリクス分析支援ツール」は、この機能も搭載している。図10に示すが、このツールはExcelがベースで、操作性とグラフ化などの可視化に優れており、効率的な分析が可能である。本ツールでは、表2の1タイプの分析が数分程度で完了する。

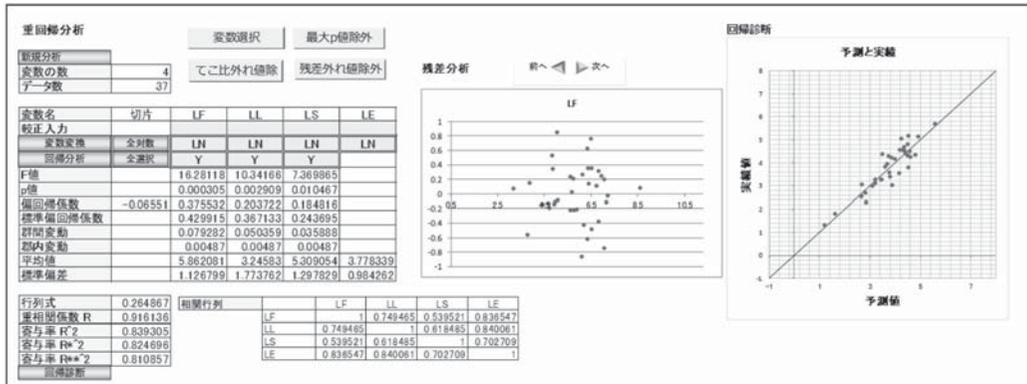


図10 マトリクス分析支援ツール（重回帰分析機能）

また、このツールは、分析の対象となる各データを対数変換する機能のほか、複雑な計算を要する回帰式の外れ値候補を自動で識別する機能もある。図11に示す「残差外れ値」と「てこ比外れ値」が瞬時に検出でき、標準工数の予測に適さない特異な性質を持つデータを簡易に除外できる。

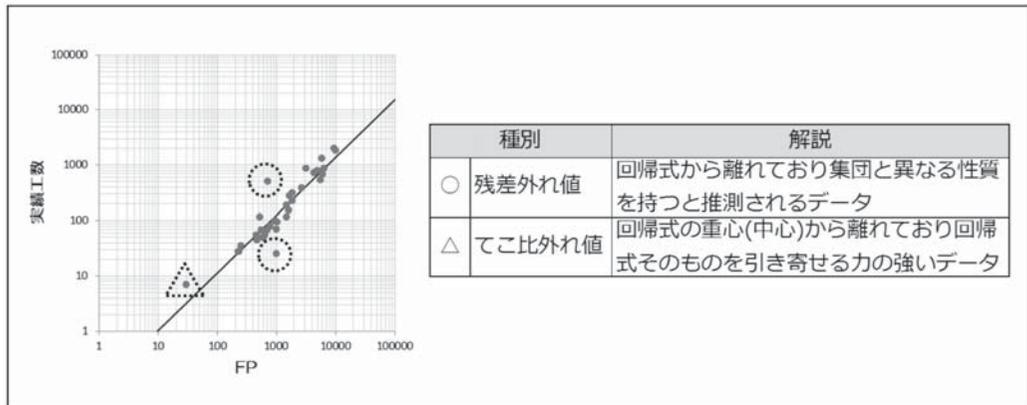


図11 マトリクス分析支援ツール（外れ値候補検出機能）

5.3 重回帰分析結果の確認

通常、回帰式の精度は主に「寄与率」という統計量で判断する。寄与率とは、FPなどの説明変数が、目的変数となる標準工数をどの程度説明できるのかを表す指標である。値は0から1の間を取り、1に近づくほど回帰式の精度が高くなる。一般的には、値が0.5以上の場合に回帰式が利用可能と判断する。

ところが、この寄与率は、説明変数を多く取り込むほど値が高くなる。例えば、データ数が

5 個の場合に、説明変数を 4 個にすると、実際には説明変数と目的変数の間に何の関係性がなくとも寄与率が 1 になる。このような回帰式は、いくら寄与率が高くても標準工数の予測には不適切である。統計学では、この誤りを防ぐため、説明変数が多くなるほどペナルティを与えて寄与率を下げる「自由度調整済み寄与率」という方式がある。

しかし今回は、この自由度調整済み寄与率を、さらに拡張した「自由度 2 重調整済み寄与率」を用いた。この寄与率は、対象データ数が少ない場合や、説明変数が多い場合のペナルティを、より大きくしたものである。開発プロジェクトの成果量となる標準工数を厳密に算出するため、この値が 0.5 以上であることを回帰式の採用の判断基準とした。

なお、この確認作業では興味深い事実も判明した。15 タイプの分析には、FP や SLOC だけで工数を予測する単回帰分析もある。しかし、単回帰分析では寄与率が 0.9 を超えるものは少なく、ほとんどが 0.6 未満の値となる。これに対し、FP や SLOC にテストケース数を加えた重回帰分析では、寄与率が 0.9 を超える回帰式が飛躍的に増加する。つまり、従来から利用されている FP や SLOC だけでの工数見積りでは、真の工数の半分程度しか予測できないのである。また、ある層別では FP と工数の関係性が強いが、同じ層別に SLOC を用いると工数との関係性がほとんど確認できないなど、層別によって利用できる規模尺度に違いがあることも判明した。

こうして、1335 パターンの回帰分析を確認し、計 167 の回帰式を標準工数の予測に利用可能と判定した。この回帰式を組み込み、標準工数と工数生産性の評価結果を算出する工数生産性評価モデルを構築した。

6. 工数生産性評価モデルによる生産性の評価結果

構築した工数生産性評価モデルを用いて、2013 年度に完了した 22 件と、2014 年度に完了した 24 件の計 46 プロジェクトの生産性を評価した。本章では、各プロジェクトと組織全体の生産性の評価結果を記述する。

6.1 各プロジェクトの評価結果

図 12 は、工数生産性評価モデルによる評価事例である。①の点線で囲んだ部分に、対象プロジェクトの FP や SLOC などの規模、実績工数 (48.4 人月)、開発タイプや主言語などのプロジェクト特性、作業対象の該当工程を入力すると、生産性の評価結果を瞬時に表示する。

②の「50.26」は成果量となる標準工数であり、③の「1.038 (= 50.26 ÷ 48.4)」が工数生産性の評価結果である。この場合、過去の同じ特性のプロジェクトでは 50.26 人月を費やす価値の開発を 48.4 人月で完了したため、3.8%の生産性向上となる。同時に④に示す「1.86 人月 (= 50.26 - 48.4)」の工数削減を達成したと評価する。

⑤は、工数生産性の評価結果を可視化したグラフである。横軸は回帰式で算出した標準工数を、縦軸は実績工数を表す。グラフ内の対角線は「標準工数 = 実績工数」で、生産性に変化がないことを示すラインである。左斜めのグレー色のエリアは「標準工数 < 実績工数」の低生産性エリアで、右斜めの白いエリアは「標準工数 > 実績工数」の高生産性エリアを示す。

黒点は、評価の対象と同じ特性を持つ過去のプロジェクトで、十字の交差する点は評価対象プロジェクトの標準工数と実績工数の関係を表す。十字の交差点がどのエリアに位置するのによって、対象プロジェクトの評価結果を視覚的に確認することができる。

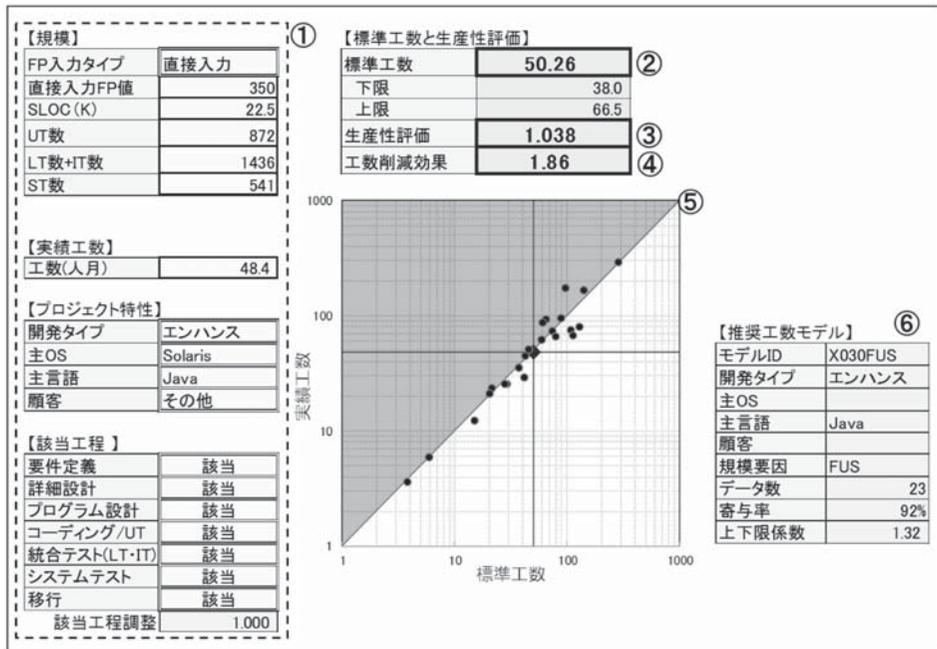


図 12 工数生産性評価モデルによる評価事例

⑥は、標準工数を算出した回帰式の情報である。この場合は、「エンハンス-Java」という特性の回帰式に、FP値(F)、単体テストケース数(U)、システムテストケース数(S)を説明変数として用いて標準工数を算出したことを示す。

このモデルで計46プロジェクトを評価した結果、ほとんどの工数生産性は、概ね0.9～1.1であり、生産性の向上と下降が±10%の幅で確認できた。また、実績工数が標準工数の上限値を上回るプロジェクトに対しては、その原因を追跡調査した。この結果、これらのプロジェクトでは、要件変更の多発、想定外の調整作業の発生、開発期間の極端な圧縮などがあり、低生産となる要因が確認できたのである。

なお、本評価と並行して、FP生産性も算出した。この結果、FP生産性では5,000FPを超える幾つかの大規模プロジェクトで、基準値の1/5程度の極端な低生産という評価が発生した。しかし、工数生産性では、同一のプロジェクトの評価が1.0を超えるという真逆のケースもあった。これは、FP生産性の問題点である規模特性の未反映を、工数生産性がカバーしているためである。

6.2 組織全体の評価結果

開発組織全体の評価は、各プロジェクトの評価結果を集計するだけである。図13は、2012年度を基準に、2013年度と2014年度の組織全体を評価した結果である。

上部のグラフで示すように、組織全体の工数生産性は、基準となる2012年度と比較し、2013年度は10.4%、2014年度は6.7%の向上であった。また、生産性向上施策の効果を判定する平均工数生産性は、2013年度が6.7%、2014年度は12.9%の向上を示しており、適用した各種の施策が有効であったと判断した。なお、工数削減効果は、2013年度が328.2人月、2014年度は116.8人月の工数削減を達成している。

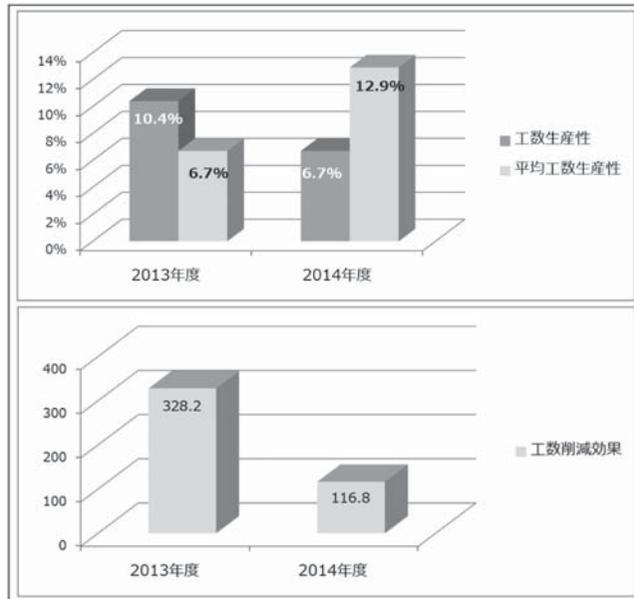


図 13 組織全体の評価結果

工数生産性の評価結果を詳細に見ると 2013 年度は、9,000FP を超える大規模な再構築プロジェクトが 20% を超える高生産性であり、この影響で工数生産性が大きく向上している。一方、2014 年度は、突出した高、又は低生産性の大規模プロジェクトが存在しなかったが、全体的にやや高生産性のプロジェクトが多い傾向であった。また、2014 年度の平均工数生産性が高い値を示しているが、これは数件の小規模プロジェクトが、極端に高い生産性であることが原因である。ただし、この数件を除いた場合でも 8% 程度の向上を確認しており、各施策の効果があつたと判断した。

7. 工数生産性評価モデルの今後の方向性

工数生産性評価モデルの最大の長所は、データの蓄積さえ行えば、あらゆる IT 企業で適用できる点である。そして、データ数の増加に伴い、予測と評価の精度がより高まる。2015 年には、最新の実績データを追加した約 300 プロジェクトのデータで、工数生産性評価モデルを更新した。データの増加で利用できる回帰式は 520 となり、2014 年に構築した初版の 167 を大きく上回る結果となった。また、多くの回帰式の寄与率が 0.9 を超えており、標準工数の予測精度が格段に高まっている。

さらに、工数生産性評価モデルの評価精度向上のため、二つの拡張を検討している。一つ目は、説明変数（規模尺度）の追加である。現行の仕組みでは、FP、SLOC、各テストケースの五つの説明変数を用いている。しかし当社は、ドキュメントのページ数や、レビュー回数などの多くのデータを蓄積しており、新たな説明変数の追加で精度の向上が期待できる。

二つ目は、プロジェクトの特性を表す定性的なデータを蓄積し、標準工数の予測に利用する仕組みである。統計学には、アンケート結果などの定性的なデータから、定量的なデータを予測する手法があり、これを適用できる。既にプロジェクトの定量的データと共に、定性的な要因を収集する仕組みも整備した。図 14 に示す「定性要因調査票」により、数値化が難しい以

下の6カテゴリ、24問の定性的要因に関する情報の収集を開始した。

- 1) 要件の状態や顧客との関係 (3 設問)
- 2) 品質・コスト・スケジュールの制約・前提条件 (4 設問)
- 3) プロジェクト・マネージャーのスキルと経験 (3 設問)
- 4) プロジェクト・メンバーのスキルと経験 (8 設問)
- 5) 開発体制や外部との関係 (3 設問)
- 6) テストや移行に関する特性 (3 設問)

1. 要件の状態や顧客との関係

1.1. 各要件は、どのような状態で顧客から提示されたか。

回答	N/A
----	-----

1. 要件は極めて明確であった
2. 要件は明確であった
3. 特に問題はなかった
4. 曖昧さや不整合があった
5. 非常に曖昧で不整合も多い内容であった

(中略)

2. 品質・コスト・スケジュールの制約・前提条件

2.1. 稼働後の品質レベルや、障害に対する顧客の制約はどのようなものであったか。

回答	N/A
----	-----

1. 品質に対する制約はまったくなかった
2. 品質に対する制約はほとんどなかった
3. 稼働後に一定の障害が発生することは許容されていた
4. 障害はほとんど許容されず、高い品質レベルを要求された
5. 障害はまったく許容されず、極めて高い品質レベルを要求された

図 14 定性要因調査票 (一部抜粋)

また、工数生産性評価モデルは、生産性評価の仕組みであると同時に、工数見積り・評価機能も包含する。標準工数の算出と工数見積りは、表裏一体の関係にあるといえる。プロジェクト早期段階における過小・過大見積りの防止を目的に、2016年4月には以下の五つの特徴を持つ「工数評価ツール」(図15)を構築し、開発プロジェクトに公開した。

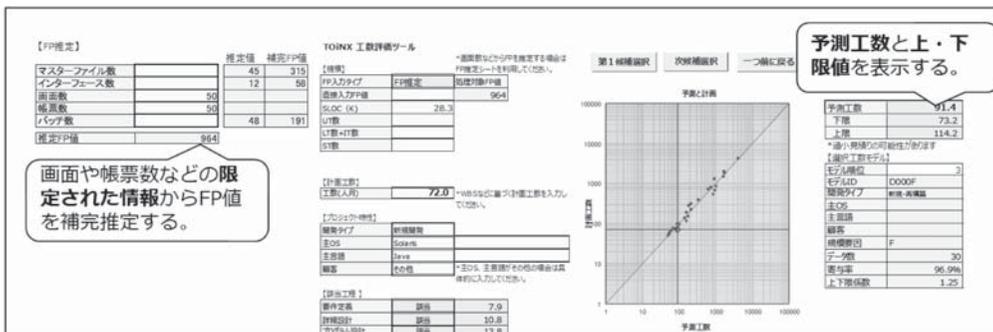


図 15 工数評価ツール

- 1) 実績データに基づく 500 超の工数予測式を内蔵
- 2) 画面数や帳票数などの限定された情報でも工数が予測可能
- 3) 工数予測式ごとの正常範囲（上下限值）を提示
- 4) 全体工程のほか、特定の工程に対する工数予測も可能
- 5) FP や SLOC の入力値が予測式の適用範囲を超える場合に警告を表示

8. おわりに

生産性を厳密に監視することで、低生産性となったプロジェクトの原因究明と再発防止の活動が開始できる。同様に、高生産性となったプロジェクトの要因を調査し、その結果を展開することで、組織全体の生産性向上を計ることができる。低コストで高品質なソフトウェアを開発し、顧客満足度を向上させるためのプロセス改善活動には、正確な生産性の計測が不可欠である。

一方で、多様な要因が複雑に絡み合う開発プロジェクトの生産性評価は、極めて難易度が高く、具体的な事例も少ない。だが、多くの産業の発展に計測が有効であったように、IT 分野でも、計測とデータ解析手法を駆使し、組織の成熟度レベルの向上を実践するときに来ている。

最後になりますが、本稿を執筆するにあたり、株式会社ワイハットの梶山昌之氏から、多大なるご支援と、ご協力を賜ったことに厚く御礼申し上げます。

-
- * 1 米国のソフトウェア工学研究所 (SEI) によって考案されたソフトウェア開発プロセスを改善するためのガイドラインで、開発組織の成熟度レベルを 5 段階で評定する。
 - * 2 FP 法の国際団体である IFPUG (International Function Point Users Group) が策定した CPM (Counting Practice Manual) と呼ばれるマニュアルに基づく計測手法。
 - * 3 多数の開発プロジェクトのデータを整理、分析した白書。独立行政法人 情報処理推進機構 (IPA) が定期的に発行している。
 - * 4 ソフトウェア開発のデータを集計・分析した結果を取りまとめたもの。一般財団法人 経済調査会が発行している。
 - * 5 SLCP-JCF (Software LifeCycle Process - Japan Common Frame) とは、ソフトウェア取引に関するガイドラインで、ソフトウェアの構想・設計から開発、保守、破棄に到る各工程の作業内容などを標準化した枠組みである。独立行政法人 情報処理推進機構 (IPA) が発行している。
 - * 6 データの分布状況の把握を目的としたグラフで、度数分布図ともいう。データの中心や、バラツキの状態などを調べるために用いる。
 - * 7 データが分析の条件を満たさない場合に、元のデータを対数にすることで条件を満たすようにするための手法。
 - * 8 複数のデータを大小の順に並べたときに中央にくる値。
 - * 9 複数のデータの中で最も個数の多い値。
 - * 10 JIS X 0141 : 2004 「ソフトウェア測定プロセス」では、情報ニーズに基づく基本測定量、導出測定量、及びその測定・計算方法などを含む広い範囲をメトリクスと定義している。本稿では主に、統計的技法を用いて設定した「基準値」と「予測モデル」の二つをメトリクスと呼ぶ。
 - * 11 対数変換の際に用いる対数関数の逆関数であり、逆対数とも呼ばれる。
 - * 12 データを同じ共通点を持つグループに分類すること、層別することで、データの特徴がより明確になることがある。
 - * 13 互いに関連のある多種類のデータの相互関連を分析する統計的技法の総称である。
 - * 14 予測の対象とする変数で、結果となる事柄に該当する。
 - * 15 目的変数の変動を説明する要因となる変数で、原因となる事柄に該当する。例えば、住宅価格が目的変数の場合は、居住面積、築年数、駅からの距離などが該当する。

- 参考文献**
- [1] 梶山昌之, 合田英二, 千野智子, 「プロジェクト初期段階のFPによる効果的見積りモデル構築手法の提案」, プロジェクトマネジメント学会 2007 年度春季研究発表大会 予稿集, (一社)プロジェクトマネジメント学会, 2007 年, P116 ~ 121
 - [2] 金子英一, 梶山昌之, 「アプリケーション・マネジメント・サービス (AMS) における開発生産性評価プロセスの確立」, IBM ProVISION Winter 2006 No.48, 日本アイ・ビー・エム(株), 2006 年, P100 ~ 107
 - [3] 佐藤浩明, 「定量的マネジメントを目指したメトリクス分析の実践~開発プロジェクトの成功確率を高めるために~」, ユニシス研究会, 2013 年
 - [4] 亀井秀敏, 佐藤浩明, 「システム開発プロジェクトの健全度可視化と定量的マネジメントへの取り組み」, IBM ユーザー研究会, 2014 年

執筆者紹介 佐藤 浩 明 (Hiroaki Sato)

1998 年東北コンピュータ・サービス株式会社 (現, 東北インフォメーション・システムズ株式会社) 入社. 2006 年より SPI (Software Process Improvement) 活動と, PMO (Project Management Office) に従事. 現在は, IT メトリクスの解析を担当している.

