

ソフトウェアプロセス改善を組織的、実証的にすすめるための

データ分析パターン言語の提案

A Proposal for a Pattern Language of Data Analysis to promote Empirical and Organizational Process Improvement

株式会社 プロセス分析ラボ
Process Analysis Laboratory Inc.
○小室 睦
○Mutsumi Komuro

早稲田大学
Waseda University
鷲崎弘宣
Hironori Washizaki

Abstract A Pattern Language of Data analysis is proposed. It includes 15 patterns which are classified into 3 categories: (1) Preparation patterns, (2) Patterns on data analysis, and (3) Patterns concerning process improvement strategies. A fundamental hypothesis about this pattern language is that it would promote empirical and organizational process improvement in software development companies. Several evaluations are conducted to verify this hypothesis, including literature survey and experimental workshop. The results of these evaluations are reported. Especially, it is shown that the pattern language has positive effect on communications among stakeholders.

1. はじめに

プロセス改善の歴史は古く、日本での源流は第2次大戦後の製造業の品質管理活動[1]に求めることができる。しかし、ソフトウェア開発にプロセス改善を適用しようとする、人的要因や環境要因の影響が大きいなど、特有の問題がある[2][3]。

品質管理活動では統計的手法を用いた工程管理が重要視されているが、ソフトウェア開発とそのプロセス改善を効果的にすすめるためにもデータ分析は有効である[4][5]。データを通じてプロセスの実施状況を客観的に把握し、改善事項の発見、改善効果の確認・評価が可能となり、さらには品質の可視化など実施状況の定量的監視・制御を実現できる[3]。さらに進んだ段階では実施結果（品質・工数・納期等）の定量的予測も可能になる[6]。また、最近一般的になりつつあるアジャイル開発においても継続的な改善は中心的な概念であり、そのためのデータを用いた可視化は重要なテーマである[7]。

しかし、有効である反面、ソフトウェア開発での定量的データの扱いには困難も伴う。上述の通り人的要因や環境要因の影響が大きい、品質管理活動での統計的手法適用では前提条件とされているデータの安定化（雑音の低減）を実現するのが難しい。さらには、人の活動（プロセス）のデータを扱うため、次のような問題が起こり得る：測定活動が形骸化して使用されていないデータが収集され続ける、人やプロジェクトの評価や叱責に測定データが使われるおそれから現場が警戒して測定活動に非協力的になる、極端な場合には虚偽の測定値が報告される[8]。

データ分析に関する文献は数多くあるが、このような適用上の実践的な問題点や対処法に触れたものは少ない。筆頭著者は SQiP 研究会 演習コース II でソフトウェア開発プロセスに関するデータ分析の講義を社会人に対して継続的に実施しているが、講義内容は分析技法、統計ツールの使用法、統計知識の解説などに偏りがちで、適用事例の解説は行うもののデータ分析を実際に適用する上での秘訣や留意点はうまく伝えられていないという問題点を感じていた。そこで、パターン言語の形でこれらの秘訣や留意点を記述することを考えた。

本論文ではソフトウェアプロセス改善を組織的、実証的にすすめるためのデータ分析パターン言語の構築、また、このパターン言語の適用可能性確認のために実施した実験結果について報告する。「ソフトウェアプロセス改善を組織的、実証的にすすめるためのデータ分析パターン言語」のことを以下では、データ分析パターン言語、もしくは単にパターン言語と呼ぶ。

2 パターン言語

2.1 パターンとパターン言語

パターンとは繰り返し現れる問題、その問題が生じる状況と解決策を一般的に適用可能な形に記述し、適切な名前を付けまとめたものである。パターン言語はいくつかのパターンをまとめて知識記述と共有を図る方法で、建築家クリストファー・アレグザンダーが住民参加のまちづくりのために、町や建物に繰り返し現れる「パターン」をまとめて言語として共有しようとしたのが始まりである[9]。このアイデアが1990年代にソフトウェア業界に取り入れられ、Gammaらのデザインパターンなどの著作[10]が生まれた。最近では、Scrum book[11]がパターン言語で記述され、また、アジャイル開発で効率的かつ効果的に品質保証を実践するための方法もパターン言語としてまとめられている[12]。

2.2 なぜパターン言語なのか

本論文ではソフトウェアプロセス改善におけるデータ分析を組織的、実証的にすすめるためにパターン言語を用いることを提案する。データ分析に関するノウハウや知識を記述する手段としてはパターン言語以外にもさまざまな手法が考えられる。パターン言語と代表的な手法との比較をするため、以下のような資料を調査し表 1 にまとめた。

ノウハウ集：ソフトウェア品質保証の肝[14]，

べからず集：Do's and Don'ts of Process Improvement[15]，データ分析の光と影[16]，

解説：品質データ分析の基本(文献[5]の 1.2 節)，

規格書：CMMI モデル[6]

表 1 他方式との比較

方式	背景/文脈	問題	解	留意点	例	関連/体系	記述の抽象性、一般性	対応する問題の検索
パターン・パターン言語	○	○	○	○	○	○	抽象的・一般的	容易
ノウハウ集	○	○	○				具体的・特殊	容易
べからず集	○	○		○			具体的・特殊	難
解説		○	○			○	抽象的・一般的	難
規格書	○	○		△	△	○	抽象的・一般的	やや難

パターンは必要な項目を網羅しており名前と関連・体系を使えば容易にその内容にたどり着ける。抽象的であることは欠点とも言えるが、一般性を持った記述になっており、広い範囲に適用できる。もともとのアレグザンダーのアイデアによれば、名前付けられたパターンを体系的に複数記述することで知識の共有やコミュニケーションに利用できることが期待できる。

本論文のパターン言語はプロセス改善のためのデータ分析に活用することを意図しており、関係者との知識共有・コミュニケーションが重要である。また、データ分析による実証的な改善を実現するためには、一般的なやり方を理解したうえで実際の適用は状況とデータに応じて自分達で判断していくことが望ましい。記述が具体的すぎるとそこに書いてある結論に飛びついてしまい、適用可能性や解の意図などを十分考えない危険性がある。以上のような考えからパターン言語による記述を採用した。

3. データ分析パターン言語の構成とその意図

3.1 想定適用範囲

データ分析パターン言語はソフトウェア開発企業内でのプロセス改善を組織的、実証的にすすめるためのデータ分析を対象としており、組織レベルで改善をすすめる EPG (改善推進者) や QA (品質保証) が使うことを想定している。しかし、以下で述べる 3 分類のうち、データ分析の準備、環境・インフラの整備に関わるパターンおよびデータ分析そのものに関するパターンは組織内の各プロジェクトでのデータ分析にも適用可能である。

3.2 分類とパターンの紹介

データ分析パターン言語ではパターンを次の 3 種類に分類して整理している：(1) データ分析のための環境・インフラの整備や準備に関わるパターン、(2) データ分析そのものに関するパターン、(3) データを活用した改善活動に関するパターン。それぞれの分類の概要は以下の通りである。

(1) データ分析の準備、環境・インフラの整備に関わるパターン

一般的にデータ分析での準備の重要性は良く知られている[5][13]。特に、ソフトウェア開発データでは人的要因、環境要因などから雑音が混入しやすいので、雑音を一つ一つ除いていく地道な作業が重要になってくる。また、測定の価値と目的を明らかにして測定活動との整合性を保証すること、データ収集・分析とそれに基づく改善が組織からの押し付けにならないよう心理的安全性を確保することもこの分類に含まれる。

(2) データ分析そのものに関するパターン

ソフトウェア開発データの分析という文脈ではデータは(開発・管理)プロセスから生み出される。データを生み出しているプロセスの実態に関心を持ち続け、どのような状態が「良いプロセス」といえるのか、また、それを明らかにする指標は何なのかという理解・洞察を確立する。そのためにはデータの意味を常に考えながらデータの示す事実で判断していくことを基本として、すすんだ段階では各プロセスの実態・実力を明らかにすること、結果に関連するプロセスの内、どれがキーとなるか理解し効果的な改善を定量的にすすめることも含まれる。

(3) データを活用した改善活動に関するパターン

有効な改善活動につなげるためには人を納得させ動かす戦略が必要である。この戦略の中には、改善のビジョンと価値を共有しながら、目標と現場の要求のバランスを取り改善の方向感を失わないこと、協力的なプロジェクトを増やし改善の輪を広げること、そして実証的に効果を確かめながら改善を着実にすすめることが含まれる。

表 2 に分類ごとのパターンの一覧を示す。なお、パターン記述内や本論文中で(他の)パターンを参照する際には<<P1 ありのままの報告>>のようにパターン ID とパターン名称を'<<'>>'で囲んで示す。

表2 パターンの一覧

分類	ID	パターン名称	問題点 (要点)	解決策 (要点)
準備・インフラ整備に関わるパターン	P1	ありのままの報告	<ul style="list-style-type: none"> 改善点・問題点があっても現場から相談がなく形式的な報告のみ 現場からの測定報告に実際の実績値どおりであるかどうか疑義がある 	<ul style="list-style-type: none"> 測定データを人やプロジェクトの評価に使用しないことを約束する。 経営層への報告には集約データを用い、個別データはあげない
	P2	価値・目的に合った測定	<ul style="list-style-type: none"> 価値を生まない測定活動 ビジネス上の目標が各プロジェクトの活動にどう反映されるのかわからない 	<ul style="list-style-type: none"> 測定内容を組織の価値・目的に整合させる <<I5 改善のビジョン・価値を示し共有する>>を適用 測定活動を見直し目的との整合性を保つ
	P3	データとばらつきを理解する	<ul style="list-style-type: none"> データに誤りや不整合が含まれる ばらつきが大きすぎて有効な分析ができない 分析結果が雑音の影響を受けやすい 	<ul style="list-style-type: none"> 運用定義を明確にする。<<P2 価値・目的に合った測定>>を適用し目的にあったメトリクス定義、測定プロセスとする 収集されたデータの概要を把握する <<A2 プロセスを考える>>を用いて、ばらつきに対するプロセス上の原因を見出し解決する 雑音がある/データ量が少ない状況ではロバストな(雑音に強い)分析手法を採用する
データ分析そのものに関するパターン	A1	データ (事実) で判断	<ul style="list-style-type: none"> 現状を正確に把握せずに主観と勘だけで判断し誤った対処や表面的な解決策しかとれない 組織の標準値との形式的比較、数値のみによる形式的なチェックが実施されている 	<ul style="list-style-type: none"> プロジェクト/チーム/組織の状態をデータから直接的に把握する データの示す状態の良し悪しの判断はあらかじめ設定しておいた判断基準を用いて行うが、<<A3 定性的な意味を考える>>も実施 問題が発生した場合にはその状況をデータで確認し対策を考える。 データが十分にある場合、問題発生にいたったプロセスを考察。プロセス上の原因の特定と再発防止を実施
	A2	プロセスを考える	<ul style="list-style-type: none"> 最終的な実績値から、結果として良かった/悪かったかはわかるが、有効な改善アクションがとれない。 結果データに対する原因分析から改善策を立てているが、改善は将来(次プロジェクト)の話であり、現状の利点は少ない 	<ul style="list-style-type: none"> 結果を生じる仕組み・メカニズムを考える。 改善目的・目標につながる「良い結果」とそれを生み出す「良いプロセス」は何かを明らかにする。 結果に対してプロセスがどう影響しているかについて仮説を立てる。見出した仮説を<<I3 1つ1つ確かめる>>にかけ、検証し、具体的な改善策へと展開
	A3	定性的な意味を考える	<ul style="list-style-type: none"> 実績値のみに拘り、改善に繋がりにくい分析、改善方向を間違えた分析をしている。 細かな分析技術や数値の差異に拘り、データの概要や意味を把握していない 	<ul style="list-style-type: none"> データ分析をする際の分析的な目的、結果の意味を確認 定性的な理解に基づき定量的な仮説を立て検証 得られた分析結果の定性的な意味を検討
	A4	プロセスの実態・実力を理解する	<ul style="list-style-type: none"> 実現不可能な目標値の設定、組織目標との形式的比較などプロジェクトの実態に合わない「定量的管理」の実施 	<ul style="list-style-type: none"> プロセス実施上の問題に対し原因分析を実施し、プロセスについての理解を深め再発防止をはかる プロセス実績データのベースラインを確立
	A5	キーとなるプロセスを見出す	<ul style="list-style-type: none"> プロセス実施の将来的な結果を予測して先手管理をしたい 改善活動は実施しているが効果がわからない 	<ul style="list-style-type: none"> 定性分析によるキープロセス候補の絞込み メトリクス収集と安定化(<<P3 データとばらつきを理解する>>) 定量分析の実施
	A6	自分たちのデータで確かめる	<ul style="list-style-type: none"> 他社、他組織の改善事例の形式的なマネをする 業界や全社の標準値を基準値としてそのまま使おうとする 	<ul style="list-style-type: none"> データ分析の際に分析の目的、分析結果の意味を確認 自分たちのデータで仮説を検証。成功事例が自部署でもうまくいくかについて分析 本格導入の前にパイロットで試行し結果を分析
	A7	一般化と特殊化のバランス	<ul style="list-style-type: none"> 組織ルールや標準値を形式的に適用した分析で実状が考慮されない 各プロジェクトの個別の事情に深入りし、一般的な分析結果が得られない 	<ul style="list-style-type: none"> <<A2 プロセスを考える>>により問題を広くとらえる。 <<A3 (定性的な) 意味を考える>>を適用 原因分析し一般化を考える
データを活用した改善活動に関するパターン	I1	(測定) フィードバックを早めに開始する	<ul style="list-style-type: none"> きちんとしたデータが揃ってから分析をしようとする分析結果をタイムリーに提供できず、プロジェクト側の満足度も低い データの精度が低いことを理由に分析を避けているが、結局、精度はあがらず分析も開始できない 	<ul style="list-style-type: none"> 定性的な分析でもよいので分析作業を開始する。フィードバック結果により測定・分析活動を改善 分析を開始して気が付いた改善点は提出したプロジェクトに連絡して改善を図る 仮説の誤りに気が付いた場合、関係者と話し合うなど必要な措置を取り、新しい仮説を立て、分析を続ける
	I2	成功事例を作り仲間を増やす	<ul style="list-style-type: none"> 改善やそれにともなう測定分析に抵抗感を持つ人やプロジェクトがいる 	<ul style="list-style-type: none"> 協力的なプロジェクトを見つけ効果を実証し、そこから適用を拡げていく
	I3	1つ1つ確かめる(仮説・検証・学習ループ)	<ul style="list-style-type: none"> 結果につながる有効な改善施策が選択・実施されていない 	<ul style="list-style-type: none"> プロセスの理解に基づき仮説を立て、判断基準を決め検証し、結果から学ぶ。この経験に基づき新たな仮説を立て、上記を繰り返す。
	I4	トップダウンとボトムアップの両面作戦	<ul style="list-style-type: none"> 問題を解決する方法が知られているのに現場には届いていない。 現場の実態に合わない標準や目標が組織から強制されている。 	<ul style="list-style-type: none"> 現場の声に耳を傾けながら、目標に向けて改善策を展開。トップダウンとボトムアップの両面から攻める。現場がメリットを感じられるように改善をすすめる
	I5	改善のビジョン・価値を示し共有する	<ul style="list-style-type: none"> データ収集を始めとする改善活動に理解・協力が得られない 組織的な目標に合わない「改善」活動が実施されている 	<ul style="list-style-type: none"> 改善のビジョン(改善結果、効果)を示す <<A6 自分たちのデータで確かめる>>を適用し、自分たちのデータで効果を示す <<I3 1つ1つ確かめる>>を適用し改善効果を確認しながら改善ビジョン実現に向けて継続的に活動

3.3 パターンの記述形式

データ分析パターン言語では以下のような項目に分けてパターンを記述している：パターン名、状況、問題、フォース、解、実施例、関連技術・関連パターン。最後の関連技術・関連パターンは選択的であり、省略可能である。表3に各記述項目の意味・意図を示す。

表3 記述項目とその意味・意図

項目名	意味・意図	項目名	意味・意図
パターン名	パターンの名称	解	問題に対する解決策
状況	パターンの適用される状況、文脈	実施例	パターンを実施した例
問題	パターンの解こうとしている問題	関連技術・関連パターン	関連する技術・統計技法、パターン
フォース	問題の解決に影響を及ぼす要因・力 パターン適用の前提条件・制約		

ページ数の制約から本論文では上記のすべての項目を含む実際のパターン記述例は示さない。興味のある読者は文献[17]を参照されたい。

4. パターンの典型的な適用順序

本パターン言語の典型的な適用順序の1例を図1に示す。

これは主に「分析に関するパターン」の適用順序を示している。中心となるのは「A4 プロセスを考える」と「I3 1つ1つ確かめる」の間のループで、これにより「良い結果を生み出す良いプロセスは何か」という問いに対する答が深められる。

分析ではまず「A1 データで判断」し、「A3 定性的な意味を考える」で意味・意図を確認するのが基本である。データ収集が進んできた段階では、「P3 データとバラツキを理解する」で変動を理解し対処しながら、各プロセスの実績の実力値を把握する。これはCMMI[6]の高成熟度改善でいうところのプロセス実績ベースラインの確立に相当する。また、「A5 キーとなるプロセスを見出す」により改善結果に寄与が大きく制御可能なプロセスを見出し、定量的管理や改善に役立つ定量的モデルを確立する。これはCMMI[6]のプロセス実績モデルに相当する。

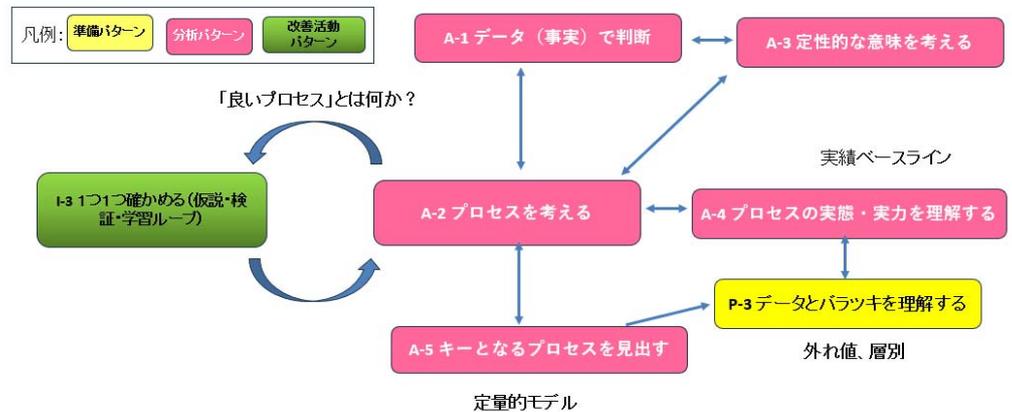


図1 パターンの適用順序例

5. 適用実験と考察

5.1 研究課題

本論文ではソフトウェアプロセス改善におけるデータ分析を組織的、実証的にすすめるためにパターン言語を用いることを提案している。この提案の妥当性を確認するために以下の3点の研究課題を設定した。

RQ1: 本パターン言語はソフトウェアプロセス改善のためのデータ分析の主要な問題点をカバーしているか？

RQ2: 本パターン言語は EPG, QA, プロジェクトリーダー, 分析担当者に対して有益な道具となりうるか？

RQ3: 組織の規模, 読者の役割 (EPG, QA, プロジェクトリーダー, 分析担当者) や経験度などに応じて本パターン言語の認知度・受容度に違いはあるか？

5.2 RQ1

RQ1については文献調査により、ソフトウェアプロセス改善のためのデータ分析に関する主要な問題点をカバーしていることを2.3節であげた5つの文献すべてについて確認した。例として、表4にソフトウェア品質保証の肝[14]と本パターン言語との対応を示す。

文献[14]ではソフトウェア品質保証に関する肝(ベストプラクティス)を108挙げており、スコープは本パターン言語よりはるかに広い。そのうち2.19品質分析・評価のマネジメントに分類されている11の肝との対応を調査した。表4よりパターン言語がこの節の肝をカバーしていることがわかるが、肝の方が個別具体的な場合があり、例えばレビューやテストに限定した肝がある。これに対し、パターンでは「プロセスを考える」というような一般的な観点か

らの記述にとどまる。これは一般的な記述とすることにより、他の状況への対応(例えば「プロセスに着目することで早い時点での品質向上活動を開始する」「プロセス上の原因を考えアクションにつなげる」)を促そうという考えに基づく。

表 4 ソフトウェア品質保証の肝と本パターン言語との関係

ID	タイトル	関係するパターン言語
【肝087】	品質向上には、計測と評価が必要	<<P-2価値・目的に合った測定>>,<<P-3データとばらつきを理解する>><<I-1 (測定) フィードバックを早めに開始する>>
【肝088】	品質指標は「良くしたいことを設定する」ことから始める	<<P-2価値・目的に合った測定>>,<<A-3 定性的な意味を考える>><<I-1 (測定) フィードバックを早めに開始する>><<I-5 改善のビジョン・価値を示し共有する>>
【肝089】	基準値を決めるからには、外れた場合のアクションを明確にする	<<A-1データ (事実) で判断>><<A-2 プロセスを考える>><<A-4 プロセスの実態・実力を理解する>>
【肝090】	品質データを集める前に目的と分析方法を明確にすること	<<P-2価値・目的に合った測定>><<I-1 (測定) フィードバックを早めに開始する>><<I-3 1つ1つ確かめる (仮説・検証・学習ループ) >>注1)
【肝091】	現場の開発プロセス把握が品質分析の前提	<<A-2 プロセスを考える>>
【肝092】	データを見誤らない	<<A-3 定性的な意味を考える>><<A-2 プロセスを考える>><<A-4 プロセスの実態・実力を理解する>>
【肝093】	データ分析を必要以上にやり過ぎないこと	<<I-1 (測定) フィードバックを早めに開始する>><<A-5 キーとなるプロセスを見出す>><<A-3 定性的な意味を考える>><<I-3 1つ1つ確かめる (仮説・検証・学習ループ) >><<A-7 一般化と特殊化のバランス>>
【肝094】	レビュー結果の評価は多面的かつ早めに行う	<<A-2 プロセスを考える>><<A-3 定性的な意味を考える>>注2)
【肝095】	開発プロセスの品質はバグが教えてくれる	<<A-2 プロセスを考える>>
【肝096】	障害票の分析項目はどのようなアクションに繋げるかを明示する	<<P-2 価値・目的に合った測定>><<A-2 プロセスを考える>><<I-5 改善のビジョン・価値を示し共有する>>注2)
【肝097】	テストの有効性はタイムリーなバグ票の分析で行う	<<A-2 プロセスを考える>><<A-4 プロセスの実態・実力を理解する>>注2)

注1) <<I5 改善のビジョン・価値を示し共有する>>にあたる組織レベルの目標との関係は触れられていない

注2) パターン言語にはテストやレビューに限定したパターンはない

5.3 RQ2

RQ2に答えるために以下のようなグループ演習形式の適用実験を実施した。

- 1 グループ 4~5名の参加者を最大2グループ集める
- シナリオを2つ (A,B) 準備する。どちらのシナリオもソフトウェア開発企業内におけるデータ収集・分析に関する問題点を複数含む
- パターン言語の知識なしで最初のシナリオに関する問題点と対策をグループで挙げてもらう。最初に時間を与えて問題点を各自で付箋に記入しその後共有してグルーピングする形式で実施
グループ1はシナリオA, グループ2はシナリオB
- パターン言語の解説を行い, (3)の結果に修正・見直しがあるか考えてもらう
- パターン言語の知識も用いて新しいシナリオの演習を実施する:グループ1はシナリオB, グループ2はシナリオA

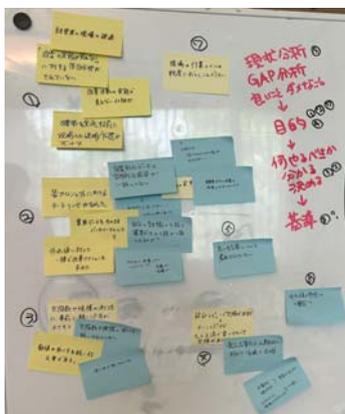


図2 演習結果(パターン未使用)



図3 演習結果 (パターン使用)

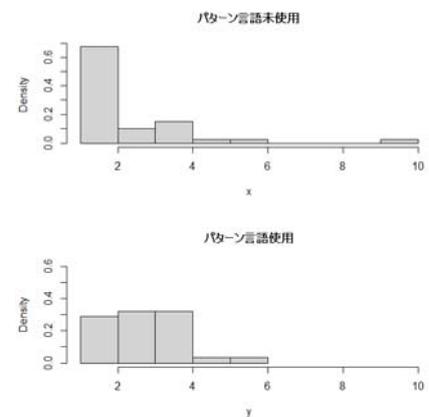


図4 グルーピングのサイズの分布

実験は3回実施した。実験参加者は21名でSQiP研究会 演習コース II の受講生9名、卒業生12名からなる。各回の参加者数は順に9名(2グループ)、4名(1グループ)、8名(2グループ)だった。最初の回は対面で模造紙と付箋を用い、残り2回はリモートでZoomとMiroで実施した。図2,図3は対面で実施した回でのパターン説明前と説明後の結果例である。実施したのは別のグループだが、評価したシナリオは同一である。パターン使用時の方が結果が整然と配置されているように見える。

演習では問題点としてあげた内容が類似した付箋をグルーピングした後に対策を考えている。各グルーピングに対しそれが含む付箋の数をそのグルーピングのサイズと定義した。図4は3回の演習の合計データについて、このサイズのヒストグラムを示したものである。この図からも‘サイズ=1’の孤立したグルーピングが減少し、各グルーピングが3,4枚のサイズにまとまっている傾向がうかがえる。この観察を定量的に明確にするために仮説検定を実施した。

検定: Wilcoxon の順位和検定 (5%有意水準)

帰無仮説: パターン言語使用・未使用に応じてサイズの分布に違いはない。

対立仮説: パターン言語使用・未使用に応じてサイズの分布は異なる。

結果: p値=0.00951, 帰無仮説を棄却し対立仮説を採択

演習の観察, アンケート結果

演習実施時の観察からパターン名称, ID が問題点の発見, 整理のコミュニケーションに使われていることを確認した。また, 模造紙, miroに残された演習結果からパターンの問題・解の記述が結果のまとめに使用されていることを確認した。

演習実施後のアンケートの「パターンを使わなかった最初の演習とパターンを使ったその後の演習でどんな違いがありましたか?」という自由記述形式の設問に対して, 全ての参加者が「議論がしやすくなった」「問題と解決策の整理ができた」など, パターンが検討に役立った旨を回答した。

なお, 実験ではパターン言語の説明後, 最初の演習結果の見直しを実施しているが, 追加・修正はあまり実施されず, いくつかの問題点と対策が追加されたのみであった。特に, 問題のグルーピングは変更されなかった。このことから, 実際のパターン言語の適用では, 最初に教育を実施して全員がパターンの内容について, 少なくとも概要を理解している状態にしておかないと, パターン言語のコミュニケーション活性化に対する効果は十分に享受できない可能性があると思われる。

5.4 RQ3

実験参加者21名にアンケート調査を実施し, 19件の回答を得た。アンケート回答者に関する基本情報は次の通りである。

データ分析初心者(受講生) 7名, データ分析経験者(卒業生) 12名

役割: QA担当12名, EPG1名, PMO 2名, PL(プロジェクトリーダー) 2名, 開発者2名

担当プロジェクトの規模: 1~10人 6名, 10~50人 10名, それ以上 3名

ライフサイクルモデル: ウォータフォール 15人, アジャイル(ハイブリッド, 混在含む) 4名

開発形態(重複回答あり): 受託開発 10名, 商用パッケージ 6名, インハウスユース 5名

図5は本パターン言語の15のパターンについて認知度・受容度を次の3段階で回答してもらった結果である:

(1) 知らなかった, こういう考え方はなかった。(2) 内容は知っていたが実施に関与/経験したことはない。(3) 実施に関与した, 経験したことがある。

分析に関するパターン A3, A4, A5 について「知らない」という回答が多かった。A4, A5 は高度な分析技法の適用も想定したパターンであり, あまり知られていないことは考え得る。しかし, <<A3 定性的な意味を考える>>は基本的なパターンであり, 6名も知らないと答えたのは意

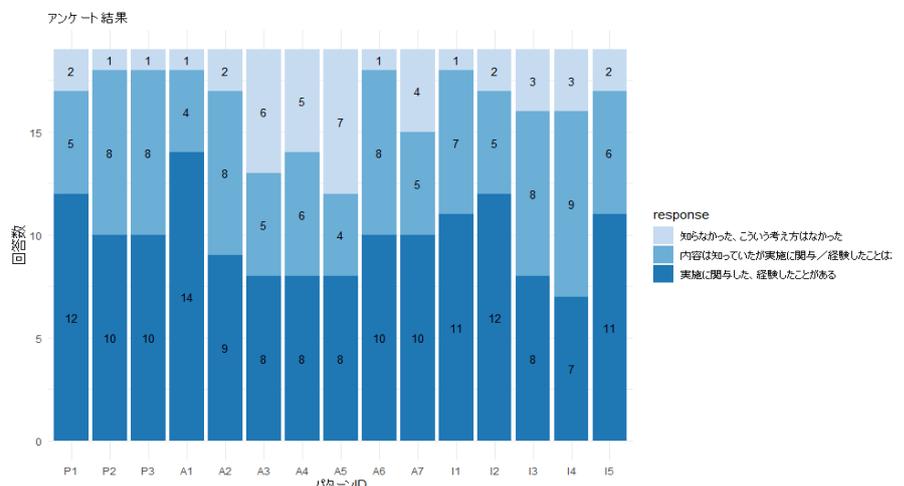


図5 認知度・受容度に対する回答結果 (全参加者)

外であった。データ分析の経験に応じて分けると、知らないと答えた人の割合は経験者では 25%だが、分析初心者では 43%とかなり高いことがわかった。

「経験がある」との回答が最も少なかったのは<<I4 トップダウンとボトムアップの両面作戦>>の 7 名であった。

この 7 名全員が組織的プロセス改善を推進する側の EPG, PMO, QA の人たちであり、一方、プロジェクト側の 4 名 (PL, 開発者) は全員「知っているが経験したことはない」という回答で「経験がある」と答えた人はいなかった。このパターンは組織レベル改善と各プロジェクトでの改善のバランスをとるといった内容であり、改善を進める側と現場との意識のズレが垣間見れたようにも思われる。

図 6 は 1~10 人の小規模プロジェクトに限定した場合の回答分布である。I4 で「経験がある」との回答は 1 件に減少しており、組織レベル改善が小規模プロジェクトまでは及びにくいことを示唆しているように思われる。I4 を除くと準備や改善のパターンについては経験度が上がっているように見える。特に P1 は 100%, I1, I2, I5 も 80%が経験ありと答えている。また、A1 も 100%になっている。これは小規模プロジェクトではボトムアップ的な (プロジェクト内の) 改善が回りやすいことを示唆しているように思われる。

以上の結果は次のような活動の必要性を示唆していると思われる：(1) データ分析の教育や実践活動において、「分析の意味を考える」ことの重要性を伝えること、(2) 組織レベル改善を進める際に現場を巻き込んで活動し、現場のプロジェクト、特に小規模プロジェクトで実施されている良い工夫を組織全体にも一般化するような活動を進めること。

5.5 妥当性への脅威

内的妥当性 パターン使用の演習は常に 2 回目に実施されており、演習のやり方になれたことで結果を整理しやすくなったという効果も考えられる。しかし、RQ2 に関しては演習中の観察、演習結果の記述内容、アンケートでの感想などからパターンがコミュニケーションに役立ったことが確認されており、上述の効果は結論に大きな影響は与えないと考える。

外的妥当性 参加者は演習コース II の受講生もしくは卒業生であり、役割でも QA が多くを占めるなど、参加者の選択に偏りがある。この点は特に RQ3 の結果分析の妥当性に影響を与え得る。今後、範囲を広げた検証を進める必要がある。

6. 関連研究、関連パターン

2.3, 6.2 で既に述べたようにパターン言語以外の形式で、データ分析に関するノウハウや知識を記述した文献が既に存在する。パターンを使用する利点は一般的、体系的な記述により問題解決およびコミュニケーションツールとしても利用できることにある。

パターンおよびパターン言語で定量的分析、プロセス改善のノウハウをまとめたものとして文献[18][19]がある。文献[18]は (CMMI モデルでいう成熟度レベル 4 にあたる) プロセス制御という定量的管理の高度な技術に関するパターン言語である。データ分析パターン言語では<<A4 プロセスの実態・実力を理解する>><<P3 データとばらつきを理解する>>がプロセス制御に直接関係するが、制御するプロセスとメトリクスを選択は<<A5 キーとなるプロセスを見出す>>、<<P2 価値・目的に合った測定>>も関係する。文献[18]のパターン言語では、コーディングやレビューといった個々の開発プロセスにまで言及しており、データ分析パターン言語より詳細である。一方、データ分析パターン言語での改善活動に関するパターンに分類されるような組織的な改善活動は明示的には論じていない。データ分析パターン言語では人との関わりやプロセスの (定性から定量への) 理解の深化も含んでおり、両パターン言語は補完関係にあると言える。

文献[19]はソフトウェアプロセス改善の理論体系を整備しようという意図のもとに大きなフレームワークを提案している。このパターン言語は CMMI, SPICE, ISO といった規格類より抽象的、一般的なレベルにあるとされ、計画、ベスト

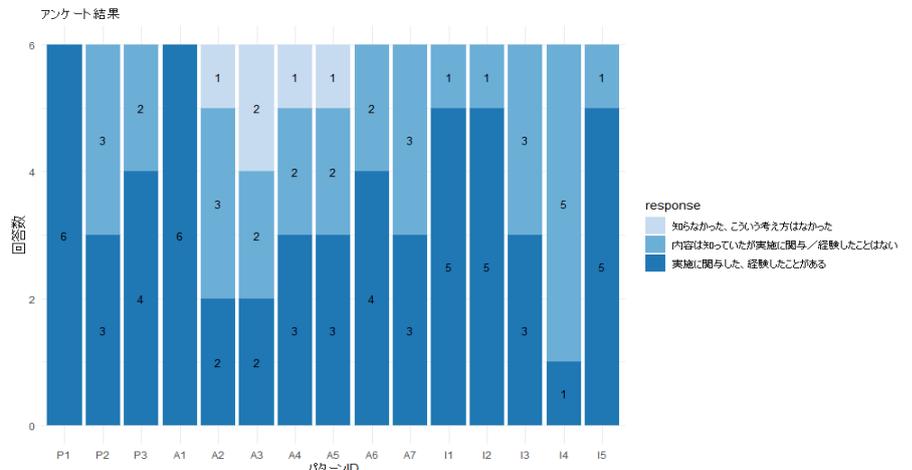


図 6 認知度・受容度に対する回答結果 (小規模プロジェクト)

プラクティス、フロー、フィードバック、システム思考、生体(リビング)システムなどのパターンからなる。データ分析パターン言語は、このパターン言語よりずっと具体的で特定のテーマに絞ったレベルを扱っている。

7. おわりに

ソフトウェアプロセス改善におけるデータ分析を組織的、実証的にすすめるためにパターン言語を用いることを提案し、3つの分類、15個のパターンからなるパターン言語を提示した。このパターン言語に対する有用性の評価を実施し、次のような結論を得た。

- (1) 本パターン言語はソフトウェアプロセス改善におけるデータ分析の問題点をほぼカバーしていることを複数の文献に対する文献調査で確認した。
- (2) 適用実験を実施し、パターン言語が問題点の整理とコミュニケーションに有用であることを確認した。
- (3) 適用実験のアンケート結果から、次のような示唆が得られたが、現状の参加者の選択には偏りがあり、さらに実験を続けて参加者の範囲を広げて検証する必要がある：<<A3 定性的な意味を考える>>の認知度が、特にデータ分析の初心者に対して低い。<<I4 トップダウンとボトムアップの両面作戦>>は現場のプロジェクトにとっては実施されている実感が少ない。小規模プロジェクトではボトムアップ的な改善の受容度・認知度が比較的に高い。

今後の展望

(1) 適用実験の範囲を広げた検証を進めたい。(2) パターン言語の普及活動をすすめ、実際の適用例を通して、改善活動や分析活動に有効であることを確認していきたい。

謝辞

実験方法の検討にご協力いただきました JASPIC SPC 分科会の皆様、データ分析勉強会の皆様、特に、麻生量嗣さん、福田秀樹さん、岩佐 賢さん、猪飼康博さん(順不同)に感謝いたします。また、実際の演習に参加いただいた SQiP 演習コース II の受講生の皆様、卒業生の皆様に感謝いたします。

参考文献

- [1] 石川 馨, 「品質管理入門」日科技連出版 (1988).
- [2] W. S. Humphrey, “Managing the Software Process” Addison-Wesley (1990).
- [3] W.A. Florac, A.D. Carleton, “Measuring the Software Process” Addison-Wesley (1999).
- [4] S.H. Kan, “Metrics and Models in Software Quality Engineering” Addison-Wesley, (2003).
- [5] 野中 誠, 小池和彦, 小室 睦, 「データ指向のソフトウェア品質マネジメント」日科技連出版 (2012).
- [6] M. B. Chrissis, M. Konrad, S. Shrum, “CMMI® Guidelines for Process Integration and Product Improvement”. Addison-Wesley (2006).
- [7] C. W. H. Davis, “Agile Metrics in Action” Manning Publications Co. (2015)
- [8] W. S. Humphrey, “Winning with Software” Addison-Wesley (2002).
- [9] C. Alexander et al., “A Pattern Language” Oxford Univ. Press (1977).
- [10] J. Gamma et al., “Design Patterns” Addison-Wesley (1994).
- [11] J. Sutherland, J.O. Coplien, The Scrum Patterns Group, “A SCRUM Book” Pragmatic Bookshelf (2019).
- [12] J. Yoder, R. Wirfs-Brock, A. Auiar, H. Washizaki, “QA to AQ Shifting from Quality Assurance to Agile Quality” PLoP 16 (2016).
- [13] 江崎貴裕, 「分析者のためのデータ解釈学入門」ソシム株式会社(2020).
- [14] 2016 SQiP ソフトウェア品質保証部長の会, 「ソフトウェア品質保証の肝」第 1.1 版
https://www.juse.or.jp/sqip/community/bucyo/file/hinshitsuoshoh_kimo_20190925.pptx
- [15] P. O’ Toole, “Do’ s and Don’ ts of Process Improvement”
https://insights.sei.cmu.edu/documents/3226/2004_017_001_23784.pdf (2004).
- [16] JASPIC SPC分科会, 「データ分析の光と影」SPI Japan 2020
https://www.jaspic.org/event/2020/SPIJapan/session4/4-4_ID019.pdf (2020).
- [17] 小室 睦, 鷺崎弘宣, 「ソフトウェアプロセス改善を組織的、実証的にすすめるためのデータ分析パターン言語(progress report)」Asian PLoP 2024 (2024).
- [18] M. V. Hilst, E. B. Fernandez, “A Pattern System of Underlying Theories for Process Improvement” PLoP 10 (2010).
- [19] J. Choma, E. M. Guerra, T.S. da Silva, “Pattern Language as Support to Software Measurement Planning for Statistical Process Control” PLoP 17 (2017).