

GQMを利用した組織目標が与えるメトリクスへの影響についての時系列解析

本田 澄^{1,a)} 鷺崎 弘宜^{2,b)} 深澤 良彰^{2,c)} 多賀 正博^{3,d)} 松崎 明^{3,e)}

概要: 企業において企業活動がゴールを達成したかどうかを評価することや、ゴールを定量的な方法で改善することは難しい。我々は GQM 法を基に定量的な方法で企業ゴールに関連する企業活動の結果を評価した。あるリリース日についてそれ以前にリリースされたプロジェクトの群と、それより後にリリースされたプロジェクトの群に分けて、二つの群について評価する。これをすべてのプロジェクトのリリース日について、継続的に統計分析を実施した。また、上期にリリースされたプロジェクトの群と下期にリリースされたプロジェクトに分けて年度内と年度をまたいで評価する定期的な統計分析を実施した。実証的事例研究として株式会社いい生活が提供する不動産業務をサポートするクラウドサービスに対して評価した。対象とするデータは3年間のプロジェクト品質評価結果である。73 プロジェクトについてリリース判定に用いている品質評価結果を利用して、継続的な統計分析と定期的な統計分析を利用し、群の変化について平均値および、分散について評価する。

1. はじめに

企業における目標を評価する枠組みとしてゴール指向モデルを用いた GQM 法があるが、評価を次の目標の改善につなげることが困難である。そこで、目標の評価を時系列データととらえ2つの方法で解析し評価する。株式会社いい生活が提供する不動産業務へ支援を行うクラウドサービスを対象とし、追加機能に対する品質に関して、品質部門の目標をもとに評価する。2016年度から2018年度にリリースされた73件のプロジェクトが対象である。分析については、それぞれのプロジェクトについてリリースの判定に利用した品質評価結果からコスト、工数、開発中に発見したバグ数、トラブル対応数、コスト当たりの指摘バグ数について時系列での変化を分析する。あるリリース日以前にリリースされたプロジェクトの群とその後にリリースされたプロジェクトの群の二群に分け、それぞれの群との間に統計的有意差があるかを検定を用いて評価する。ま

た、2016年度から2018年度の3年間について各年度の上期にリリースされたプロジェクトと下期にリリースされたプロジェクトの群に分け、合計6の群について前後の群との間に統計的有意差があるかを検定を用いて評価する。結果として、企業目標に掲げる目標値について優位な差があることがわかった。

2. 背景

GQM法はメトリクスと目標との対応関係を明確にするために、Basiliらによって提案された[4]。目標となるGoalと、その目標を達成できたかどうかを評価するQuestionと、Questionに対する評価対象となるMetricsについて階層的なモデルを構築し、それぞれの対応関係を明確にできる。

GQM法は目標と質問とメトリクスを明確にし、開発者やマネージャにメトリクスが何を意味し、どのように計測されるのかの解釈を与える。そして、プロジェクトが何を達成できて、何を達成できなかったかを示す。複数の研究者は企業のプロジェクトに対してGQM法を適用し、開発中のプロジェクトの結果を分析している。

企業は環境の変化に対応するために、企業の目標を変更しながら製品やサービスを継続的に開発している。このような目標の変化は企業のそれぞれのプロジェクトの開発目標のみならず、製品やサービスの品質への影響も考えられる。つまり、将来の目標を考えるためには、目標の変更が

¹ 大阪工業大学
Osaka Institute of Technology, Hirakata-City, Osaka 573-0196, Japan
² 早稲田大学
Waseda University, Shinjuku, Tokyo 169-8555, Japan
³ 株式会社いい生活
e-Seikatsu Co., Ltd., Minato, Tokyo 106-0047, Japan
a) kiyoshi.honda@oit.ac.jp
b) washizaki@waseda.jp
c) fukazawa@waseda.jp
d) masahiro.taga@e-seikatsu.co.jp
e) akira.matsuzaki@e-seikatsu.co.jp

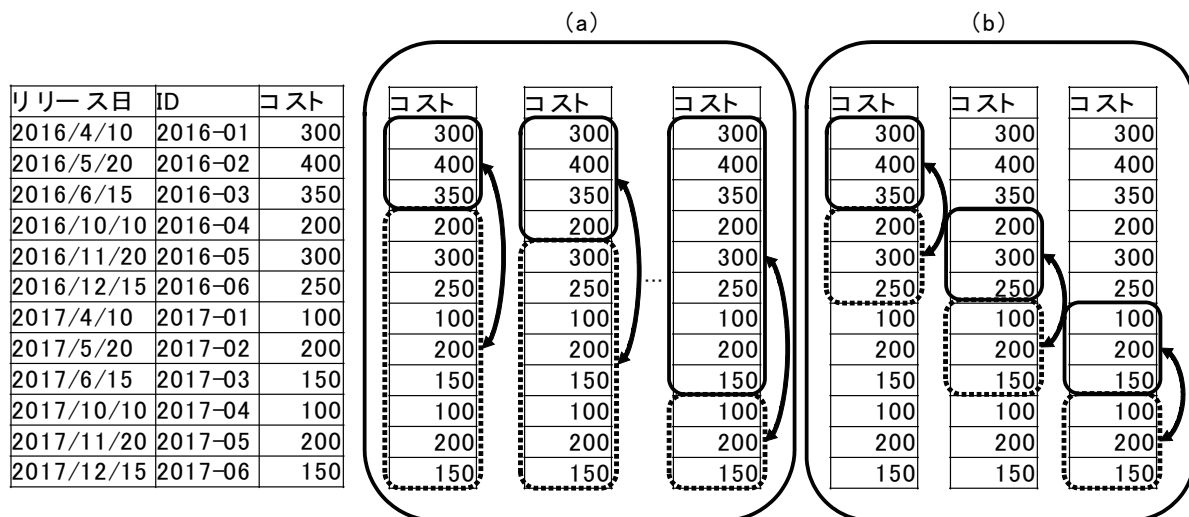


図 1 分析方法

メトリクスに与える影響を分析し、それらの関係性を理解することが重要である。

また、我々の研究グループではGQM法を複数の企業において適用しゴールの評価や改善の提案を行っている。NakaiらはCIツールを利用した開発中のプロジェクトのメトリクスを可視化する方法を提案している[1]。Tsudaらはソフトウェアの品質をISO/IEC 25000 Systems and Software Quality Requirements and Evaluation (SQuaRE) シリーズ[2]の品質メトリクスについて、GQM法に基づいた評価の枠組みを構築し21の商業ソフトウェア製品に適用している[3]。

3. 分析方法

プロジェクトの品質評価結果を時系列データととらえ2つの方法で解析し評価する。あるリリース日以前にリリースされたプロジェクトの群とその後リリースされたプロジェクトの群の二群に分け、それぞれの群との間に統計的有意差があるかを検定を用いて評価する。これにより、目標を立てた際に、どの時点でメトリクスが変化したのかを理解できると考えられる。

また、2016年度から2018年度の3年間について各年度の上期にリリースされたプロジェクトと下期にリリースされたプロジェクトの群に分け、合計6の群について前後の群との間に統計的有意差があるかを検定を用いて評価する。これにより、どの期間でメトリクスが変化したのかを理解できると考えられる。

3.1 継続的な統計分析

図1には提案手法のコストについての例を示す。まず、それぞれのプロジェクトのIDと対応する、リリース日とコストに関する情報を収集する。

図1の(a)には継続的な評価方法の例を示す。あるリリース日以前にリリースされたプロジェクトの群と、それより後にリリースされた群の2つの群に分ける。この二群についての平均値に有意な差があるかどうかを評価する。評価においては、ウェルチのt検定を用いて、有意水準として0.05として評価する。

図1の(a)内の左に示す、実線で囲った群と点線で囲った群を用いて評価し、次に、図1の(a)内の中央に示す、実線で囲った群と点線で囲った群を用いて評価する。一つずつ群から増減させ各リリース日について評価する。これをすべてのプロジェクトの各リリース日について各群に分け、それぞれの群に対して検定する。

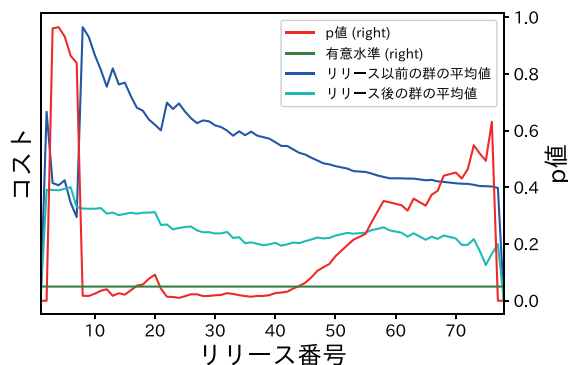


図 2 継続的な統計分析

図 2 に、継続的な統計分析の結果を示す。簡単のために横軸にリリースの順番を示すリリース番号を示す。左の縦軸に対応するメトリクスの値を示す。実際値については守秘義務のため表示していない。右の縦軸に p 値を示す。赤い実線は p 値を示し、緑の実線は p 値の有意水準 (0.05) を示す。青い実線はあるリリース番号以前のプロジェクトの群の平均値を示し、水色の実線はあるリリース番号の後のプロジェクトの群の平均値を示す。

3.2 定期的な統計分析

図 1 の (b) には定期的な評価方法の例を示す。6 か月を一つの期間とし、前後の 6 か月についての平均値に有意な差があるかどうかを評価する。評価においては、ウェルチの t 検定を用いて、有意水準として 0.05 として評価する。

図 1 の (b) 内の左に示す、実線で囲った初めの 6 か月の群と点線で囲った次の 6 か月の群を用いて評価し、その後、図 1 の (a) 内の中央に示す、実線で囲った群と点線で囲った群を用いて評価する。

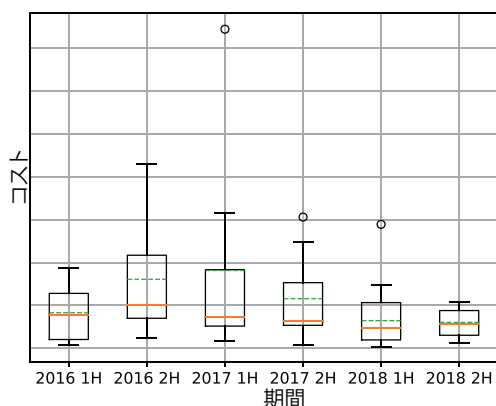


図 3 定期的な統計分析

図 3 に、定期的な統計分析の評価に利用する箱ひげ図を示す。横軸はそれぞれの期間を示す。縦軸は対応するメトリクスの値を示す。オレンジの実線は群の中央値を示し、緑の実線は群の平均値を示す。

4. 評価と結果

株式会社いい生活が提供する不動産業務へ支援を行うクラウドサービスを対象とし、追加機能に対する品質に関して、品質部門の目標をもとに評価する。2016 年度から 2018 年度にリリースされた 73 件のプロジェクトが対象である。分析については、それぞれのプロジェクトについてリリースの判定に利用した品質評価結果からコスト、工数、開発中に発見したバグ数、トラブル対応数、コスト当たりの指摘バグ数について時系列での変化を分析する。本論文で取り扱う工数はテスト工程における工数である。

表 1 に対象とする品質部門における GQM を示す。対象の品質部門では 2016 年度から 3 年の方針として品質の向上を目標として改善活動を行っている。

4.1 評価実験と結果

継続的な統計分析と定期的な統計分析の 2 種類を実施した。継続的な統計分析では、あるリリース日以前にリリースされたプロジェクトの群とその後リリースされたプロジェクトの群の二群に分け、それぞれの群との間に統計的有意差があるかを検定を用い、p 値が有意水準より下回る箇所があるかどうかを評価した。定期的な統計分析では、2016 年度から 2018 年度の 3 年間について各年度の上期にリリースされたプロジェクトと下期にリリースされたプロジェクトの群に分け、合計 6 の群について前後の群との間に統計的有意差があるかを検定を用い、p 値が有意水準より下回る期間があるかどうかを評価した。

4.1.1 継続的な統計分析の結果

図 4-8 にコスト、工数、開発中に発見したバグ数、トラブル対応数、コスト当たりの指摘バグ数についての継続的な統計分析による結果を示す。それぞれの図の軸、および凡例は図 2 と同じである。

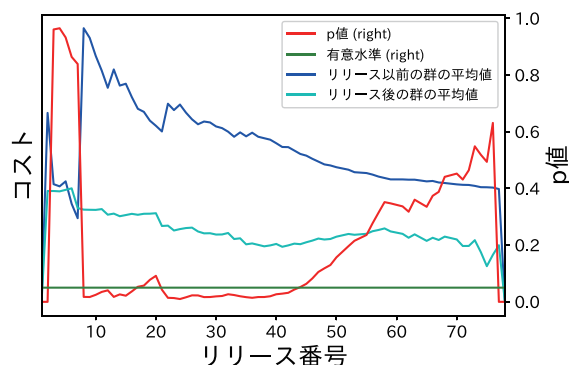


図 4 コスト

図 4 から、コストについておおよそ中央付近では p 値が有意水準より下回っており、変化していることがわかる。また、前後の平均値については前の平均値の方が低いことがわかる。

図 5 から、工数について、おおよそ全体的に p 値が有意水準より下回っており、変化していることがわかる。また、前後の平均値については前の平均値の方が低いことがわかる。

図 6 から、開発中に発見されたバグ数について、おおよそ中央付近では p 値が有意水準より下回っており、変化していることがわかる。また、前後の平均値については前の平均値の方が低いことがわかる。

図 7 から、トラブル対応数について、おおよそ中央付近では p 値が有意水準より下回っており、変化していること

表 1 GQM モデル

Top Goal	Goal	Question	Metrics
高い品質	発見されたバグが少ない	開発中に発見されているバグは少ないか?	開発中に発見されたバグ数
		リリース後のトラブルは少ないか?	リリース後のトラブル数
	効率的なテストができている	コストを抑えられているか?	コスト
		テスト工数を抑えられているか?	テスト工数
		コスト当たりの指摘バグ数は十分か?	コストあたりの指摘バグ数

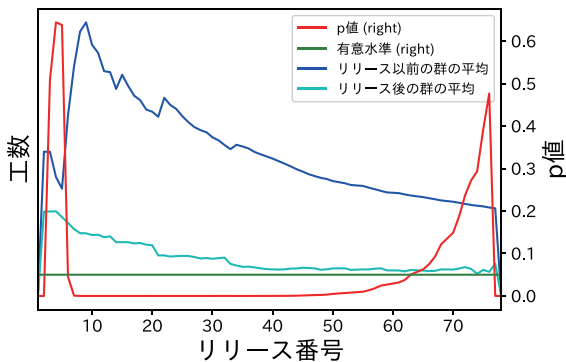


図 5 工数

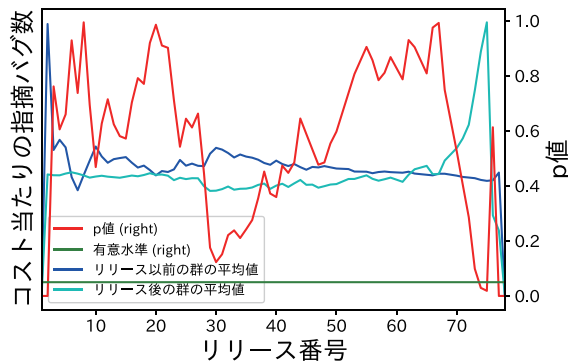


図 8 コスト当たりの指摘バグ数

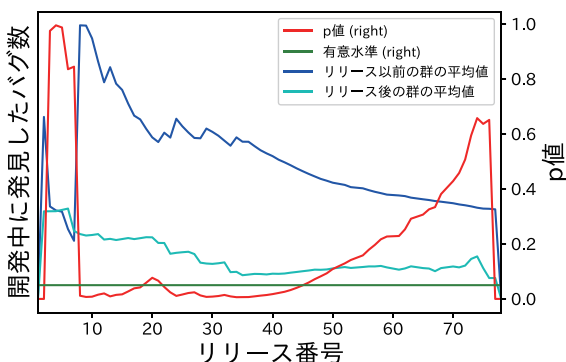


図 6 開発中に発見されたバグ数

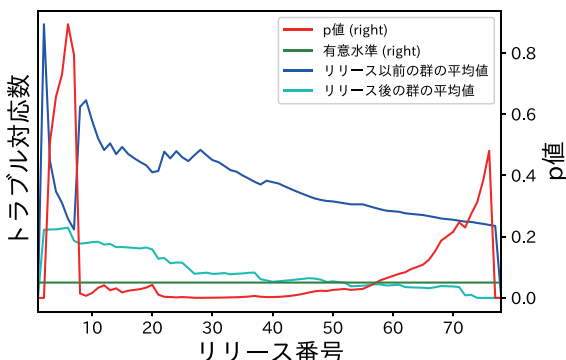


図 7 トラブル対応数

がわかる。また、前後の平均値については前の平均値の方が低いことがわかる。

図 8 から、コスト当たりの指摘数について、おおよそ全体的に p 値は有意水準より上回っており、特段の変化がな

いことがわかる。

4.1.2 定期的な統計分析の結果

2016 年度上期から、2018 年度下期までの 6 の群に分け、それぞれを前後の群と比較した。比較した対象は表 2 に示す 5 点である。

表 2 対象とする期間

記号	対応する期間
①	2016 年度上期と 2016 年度下期
②	2016 年度下期と 2017 年度上期
③	2017 年度上期と 2017 年度下期
④	2017 年度下期と 2018 年度上期
⑤	2018 年度上期と 2018 年度下期

図 9-13 にコスト、工数、開発中に発見したバグ数、トラブル対応数、コスト当たりの指摘バグ数についての定期的な統計分析で用いた箱ひげ図を示す。それぞれの図の軸、および凡例は図 3 と同じである。1H は上期を意味し、2H は下期を意味する。2016 1H は 2016 年度上期を意味する。

図 9 から、コストについて、2016 年度上期から 2017 年度の上期まで平均値が増加していることがわかる。また、2017 年度の上期から 2018 年度の下期について減少していることがわかる。

図 10 から、工数について、2016 年度上期から 2018 年度の下期まで平均値が減少していることがわかる。また、分散についても減少していることがわかる。

図 11 から、開発中に発見したバグ数について、2016 年度上期から 2017 年度の上期まで平均値が増加していることがわかる。また、2017 年度の上期から 2018 年度の下期について減少していることがわかる。加えて、分散につい

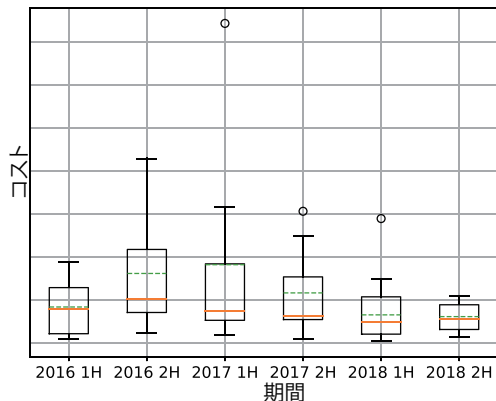


図 9 コストに関する箱ひげ図

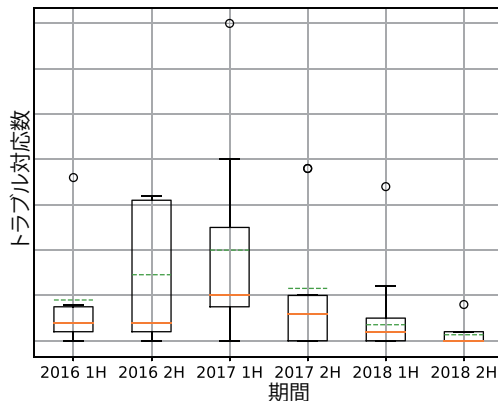


図 12 トラブル対応数に関する箱ひげ図

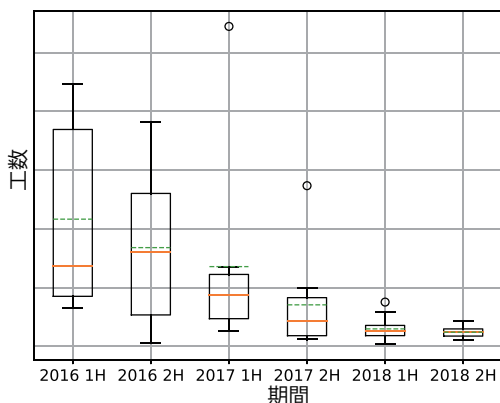


図 10 工数に関する箱ひげ図

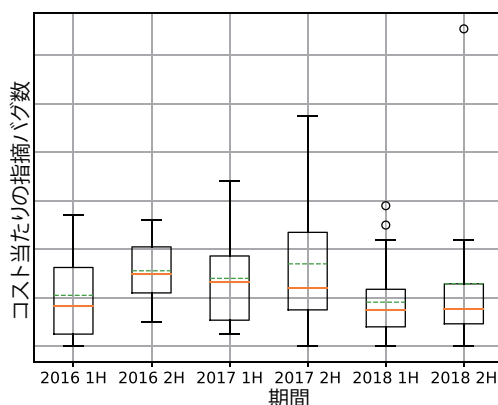


図 13 コスト当たりの指摘バグ数に関する箱ひげ図

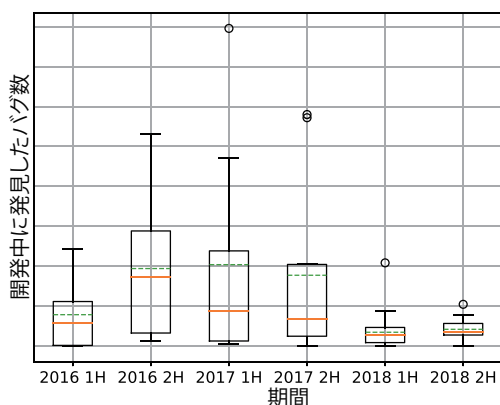


図 11 開発中に発見したバグ数に関する箱ひげ図

ては 2017 年度下以降については減少していることがわかる。

図 12 から、トラブル対応数について、2016 年度上期から 2017 年度上期まで平均値が増加していることがわかる。また、2017 年度上期から 2018 年度下期について減少していることがわかる。加えて、分散については 2017 年度下以降については減少していることがわかる。

図 13 から、コスト当たりの指摘バグ数について、2016

年度上期から 2018 年度下期について特段の変化は見られないことがわかる。

図 9-13 に示すメトリクスについて、表 3 に評価結果の p 値をそれぞれ示す。コスト、工数、開発中に発見したバグ数、トラブル対応数、コスト当たりの指摘バグ数を評価対象とし、有意水準 0.05 以下であるものについては太字として表示した。

対象としたメトリクスについて、時期が後になるほど減少する傾向があった。また、工数、開発中に発見したバグ数、コスト当たりの指摘バグ数については 2017 年度下期と 2018 年度上期で統計的有意差があることがわかる。

4.1.3 分散に関する分析

最後に、それぞれのメトリクスについて分散の変化を評価した。各メトリクスの 2016 年度上期の分散の値を 1 として、それぞれの期について表したものを表に示す。

コスト、開発中に発見したバグ数、コスト当たりの指摘バグ数については 2016 年度上期から 2017 年度上期まで増加しているが、2017 年度下期から 2018 年度下期まで減少している。工数については、減少していることがわかる。トラブル対応数については、あまり傾向がみられないこと

表 3 期間についての p 値

	①	②	③	④	⑤
コスト	0.248157	0.850004	0.474178	0.055927	0.823278
工数	0.600721	0.69897	0.32108	0.007852	0.305702
開発中に発見したバグ数	0.21663	0.940239	0.835352	0.001029	0.560797
トラブル対応数	0.785332	0.247303	0.569032	0.327211	0.257778
コスト当たりの指摘バグ数	0.338087	0.748916	0.643241	0.031223	0.303093

表 4 期間についての分散

	2016-1H	2016-2H	2017-1H	2017-2H	2018-1H	2018-2H
コスト	1.000000	3.814082	11.582332	2.035315	0.667859	0.218681
工数	1.000000	0.626099	0.877189	0.205671	0.009580	0.003068
開発中に発見したバグ数	1.000000	0.547294	0.987013	2.044318	0.532552	2.848278
トラブル対応数	1.000000	1.348688	3.112356	1.311426	0.212849	0.029049
コスト当たりの指摘バグ数	1.000000	4.050481	8.553260	5.818850	0.167137	0.098599

がわかる。

4.2 考察

継続的な統計分析と定期的な統計分析の結果について、それぞれの違いを表 5 にまとめる。継続的な統計分析では、コスト当たりの指摘数を除いて中央付近での統計的有意差がある時点があったが、定期的な統計分析においてはコストおよびトラブル対応数を除いて有意差があった。

4.2.1 継続的な統計分析に関する考察

継続的な統計分析では、コスト、工数、開発中のバグ数、トラブル対応数について平均値の差について統計的有意差があることが分かった。これは表 1 に示す、Goal“発見されたバグが少ない”については、継続的な変化が見られたことを示す。単純な時系列解析を適用することで、メトリクスの値が変化した時点を理解することができた。

一方で、前後の群に母集団の数の違いが顕著に表れていることがわかる。評価時点が全体のうちの初めや終わり付近では、評価対象となる母集団の横軸の端においては、数が極めて少ない群と、それに比べて多い群での評価となり、比較の有効性を考えなければならない。

4.2.2 定期的な統計分析に関する考察

定期的な統計分析では、工数、開発中のバグ数、コスト当たりの指摘バグ数について平均値の差について統計的有意差があることが分かった。継続的な統計分析と定期的な統計分析の結果を合わせて、表 1 を評価すると Goal“発見されたバグが少ない”および、Goal“効率的なテストができてきている”については、変化が見られたことを示す。

コストおよびトラブル対応数については、継続的な統計分析では有意差があると示しているが、定期的な統計分析では有意差がないと示している。一方で、コスト当たりの指摘バグ数については、継続的な統計分析では有意差がないと示しているが、定期的な統計分析では有意差があると示している。これは、継続的な統計分析と定期的な統計分析における比較した群の違いが関係していると考えられ

る。加えて、本研究では上期と下期のように 6 か月に分けて評価したが、この期間を変更することによって評価結果が変わることも考えられる。

4.2.3 分散の評価に関する考察

コスト、工数、トラブル対応数、開発中に発見したバグ数については、2017 年度上期から分散は減少している。マネージャへのインタビューにおいて、アジャイル開発への取り組みが影響したと考えられると発言があった。各プロジェクトの規模を小さくし、同程度にする目標を達成するために実施した施策のため、分散が減少したと考えられる。コスト当たりの指摘バグ数については、プロジェクトの規模の縮小に伴いコストと指摘バグ数の両方に影響を与え、分散の変化が顕著に表れなかったと考えられる。

4.3 制限

それぞれのプロジェクトは異なるメトリクスを持ち異なる機能を実現しているため、評価において妥当性への脅威となると考えられる。ただし、本論文では一つのクラウドサービスについて評価しているため大きな影響とは考えていない。本論文では 3 年間のプロジェクトを評価しており、市場環境による要因で複数のメトリクスに影響があったかもしれない。プロジェクトマネージャに対するインタビューにおいて、それぞれのプロジェクトをよくするために、開発スタイルをアジャイル開発へ変更したことが挙げられた。このような変更は、評価したメトリクスに影響を与えたと考えられる。こうした要因が与えた影響も認識する助けとなったことも挙げられた。

5. 関連研究

Monden らは GQM 法を利用したプロジェクトモニタリングの方法を提案している [5]。彼らの手法は開発中のプロジェクトに対して、開発者の考えに合うように変更できる GQM モデルを複数提供し、可視化できる手法を提案している。我々の論文では組織の目標について時系列解析の

表 5 評価結果

	継続的な統計分析	定期的な統計分析
コスト	中央付近での有意差あり	有意差なし
工数	中央付近での有意差あり	有意差あり
開発中に発見したバグ数	中央付近での有意差あり	有意差あり
トラブル対応数	中央付近での有意差あり	有意差なし
コスト当たりの指摘バグ数	中央付近での有意差なし	有意差あり

手法を適用して評価した。

Boerman らは、メトリクスをモニタリングすることによって、アジャイル開発の状況を測定しモニタリングする GQM 法に基づいた手法を提案している [6]。彼らはプロダクトバックログを集め、それぞれのスプリントで評価し、メトリクスの傾向を評価した。彼らは統計分析について言及していないが我々の論文では統計分析を用いて企業目標の評価を行った。

Lavazza らは、GQM 法を用いた KPI の管理プロセスを提案している [7]。彼らは UML 図を利用して系統的に KPI を決定するプロセスを構築し、MUSES プロジェクトの KPI の決定に適用した。彼らはいくつかのメトリクスの可視化方法を示しているが、メトリクスの評価において統計的な方法は利用していない。

6. まとめと展望

企業目標の評価について、GQM 法を利用しメトリクスを定め、収集したメトリクスを時系列データととらえ 2 つの方法で解析し評価した。株式会社いい生活が提供する不動産業へ支援を行うクラウドサービスを対象とし、追加機能に対する品質に関して、品質部門の目標をもとに評価した。2016 年度から 2018 年度にリリースされた 73 件のプロジェクトが対象とし、あるリリース日以前にリリースされたプロジェクトの群とその後リリースされたプロジェクトの群の二群に分け、それぞれの群との間に統計的有意差があるかを検定を用いて評価する継続的な統計分析を実施した。また、2016 年度から 2018 年度の 3 年間について各年度の上期にリリースされたプロジェクトと下期にリリースされたプロジェクトの群に分け、合計 6 の群について前後の群との間に統計的有意差があるかを検定を用いて評価する定期的な統計分析を実施した。結果として、二つの分析を実施することで、企業目標に掲げる目標値について優位な差があることがわかった。

今後は企業目標がメトリクスに与える影響の時間的な遅延の評価を実施する。影響を与える時間を理解することで、今後の目標の策定に明確な時期の設定などに役立つと考えられる。また、定期的な統計分析においては、上期と下期という 6 か月単位で分析したが、1 か月や 1 年などの単位で分析することでどのような影響が出るかも評価したい。

謝辞 本研究は JSPS 科研費 JP19K20242 の助成を受け

たものです。

参考文献

- [1] Nakai, H., Honda, K., Washizaki, H., Fukazawa, Y., Asoh, K., Takahashi, K., Ogawa, K., Mori, M., Hino, T., Hayakawa, Y., Tanaka, Y., Yamada, S. and Miyazaki, D.: Initial Industrial Experience of GQM-Based Product-Focused Project Monitoring with Trend Patterns, *Software Engineering Conference (APSEC), 2014 21st Asia-Pacific*, Vol. 2, pp. 43–46 (online), DOI: 10.1109/APSEC.2014.91 (2014).
- [2] ISO/IEC: ISO/IEC 25000:2014 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE, *Tech. Rep.*, ISO/IEC (2014).
- [3] Tsuda, N., Washizaki, H., Honda, K., Nakai, H., Fukazawa, Y., Azuma, M., Komiyama, T., Nakano, T., Suzuki, H., Morita, S., Kojima, K. and Hando, A.: WSQF: Comprehensive Software Quality Evaluation Framework and Benchmark Based on SQuaRE, *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*, ICSE-SEIP '10, Piscataway, NJ, USA, IEEE Press, pp. 312–321 (online), DOI: 10.1109/ICSE-SEIP.2019.00045 (2019).
- [4] Basili, V. R., Caldiera, G. and Rombach, H. D.: The Goal Question Metric Approach, *Encyclopedia of Software Engineering*, John Wiley & Sons, Inc, pp. 528–532 (1994).
- [5] Monden, A., Matsumura, T., Barker, M., Torii, K. and Basili, V. R.: Customizing GQM Models for Software Project Monitoring, *IEICE Transactions on Information and Systems*, Vol. E95.D, No. 9, pp. 2169–2182 (online), DOI: 10.1587/transinf.E95.D.2169 (2012).
- [6] Boerman, M. P., Lubsen, Z., Tamburri, D. A. and Visser, J.: Measuring and Monitoring Agile Development Status, *Proceedings of the 2015 IEEE/ACM 6th International Workshop on Emerging Trends in Software Metrics*, WETSOM '15, Washington, DC, USA, IEEE Computer Society, pp. 54–62 (online), DOI: 10.1109/WET-SoM.2015.15 (2015).
- [7] Lavazza, L., Frumento, E. and Mazza, R.: GQM-Based Definition and Evaluation of Software Project Success Indicators, *Software Technologies* (Lorenz, P., Cardoso, J., Maciaszek, L. A. and van Sinderen, M., eds.), Cham, Springer International Publishing, pp. 228–249 (2016).