

エンタープライズ系ソフトウェアの改良開発時の信頼性に影響を与える質的要因の分析

古山恒夫^{1,a)}

概要：IPA が収集したエンタープライズ系ソフトウェアプロジェクトデータを分析することにより改良開発プロジェクトにおける信頼性、具体的には稼働後不具合数に影響を与える質的要因を明らかにした。稼働後不具合数がゼロの値をもつプロジェクトが全体の 43% を占めるため、すべての変数を対数化して線形回帰分析を行うという方法は適用できない。そこで経済学や社会学などで用いられている負の二項回帰モデルを用いて分析した。分析の結果、改良開発プロジェクトにおいて信頼性を向上させるためには、新規開発プロジェクトの場合と同様にユーザ側の協力が重要な役割を果たすことがわかった。すなわち、ユーザ担当者の要求仕様への関与、ユーザ担当者の豊富なシステム経験と設計内容の十分な理解が信頼性を向上させる。また、デバッグ・テストツールの利用も信頼性向上に重要な役割を果たすことがわかった。改良開発プロジェクトは新規開発プロジェクトよりも信頼性が高いが、その理由のひとつが新規開発プロジェクトに比べてユーザ側の関与やスキルレベル、開発担当者のスキルレベルという点でレベルが高いすなわち開発プロセスが成熟していることも明らかにした。これらの分析結果は改良開発プロジェクトに関する信頼性向上の視点からのプロセス改善に役立てることができる。

Analysis of Qualitative Factors that Affect Reliability of Modified Enterprise Software

1. はじめに

これまでのソフトウェアの信頼性に関する研究を目的という観点からみると、大きく次の 4 つに分類できると考えられる。

(a) 出荷時の製品に含まれる欠陥^{*1}数の予測

テスト工程時の欠陥検出過程から残存欠陥数を予測する、ソフトウェア信頼度成長モデル (SRGM) が研究の代表的なものである。

(b) テスト工程時の欠陥数の予測

これはテスト工数を早期に把握するのに有効である。

(c) 欠陥を含む可能性の高い (fault-prone) モジュールの識別

この分野の研究は最近の信頼性に関する研究で主流となっているものである。Fault-prone モジュールが識別さ

れれば、そこにテスト資源を集中することにより、効率よくテストを行うことができる。

(d) 稼働後の信頼性に影響を与える要因の抽出

この分野の研究は (a)~(c) までの研究と異なり、欠陥数の予測までは踏み込んでいない。

本論文は (d) の分野に属するものであるが、この分野の研究はまだ少なく、筆者の知る限り 5 件しかない ([1], [2], [3], [4], [5])。このうち [1], [2], [3] の研究は同一の指導者によって同じ方法で実施されたという点でひとつの研究であると言える。

これらの研究では、さまざまなレベルの開発担当者が 32 の「環境要因」と呼ばれる影響要因の候補に対して信頼性^{*2}への影響の程度を 8 レベルで評価することにより影響度を推定し、テスト網羅率やプログラム仕様の変更頻度が信頼性に影響を与えると思われることを示している。

文献 [4] の研究では目的変数として 5 つの品質特性を取り上げていてその中のひとつに信頼性がある。説明変数として、組織的な変数を 10 個、個人的な変数を 5 個、技術的な

¹ 東海大学
Tokai University, Kitakaname 4-1-1, Hiratsuka City, Kanagawa, 259-1292, Japan

^{a)} furuyama@tokai-u.jp

^{*1} 本論文では特に断らない限り、障害の同義語として使用する。

^{*2} 論文では具体的な定義は行われていない。

変数を9個とりあげている。情報部門の役員 (Executives) が目的変数と説明変数に関する質問に対して5段階評価の回答を行っている。分析の結果、特に情報部門の管理者 (director) のランクおよびユーザからの仕様変更要求回数 が信頼性に大きな影響を与える要因であることを明らかにした。

これらの分析結果はいずれも興味深いものであるが、これらの研究における信頼性の評価は開発担当者の経験に基づいた主観的なものであり、例えば稼働後に発見された欠陥数などの客観的なデータで評価しているわけではない。

これに対して文献 [5] では、IPA 社会基盤センター^{*3}が2004年から収集を始めた4,000件を超えるエンタープライズ系プロジェクトデータを分析して新規開発プロジェクトにおける稼働後のソフトウェアの信頼性に影響を与える要因を明らかにした。収集項目の中には、システム稼働後 (サービスイン後) のソフトウェアの信頼性を客観的に評価できる発生不具合数、すなわち発生不具合現象数と呼ばれる故障発生数と発生不具合原因数と呼ばれる検出障害数、およびこれら発生不具合数に影響を与えると思われる多くのプロセス変数が含まれる。

本論文は、文献 [5] が新規開発プロジェクトに対して分析したのに対して、それと全く同じ分析方法を用いて、改良開発プロジェクトを対象として分析した結果を報告するものである。すなわち、改良開発プロジェクトにおける信頼性 (システム稼働後の発生不具合数) に大きな影響を与える質的変数を明らかにしたものである。

2章で分析対象データを紹介する。3章で分析方法を示す。4章で分析結果を示す。5章で分析結果に対する考察を、6章でまとめを述べる。

文献 [5] と本論文の違いは、開発種別が新規開発プロジェクトか改良開発プロジェクトかだけなので、1章から3章までは、文献 [5] から本論文だけで閉じる必要のある部分を抜粋して記載している。

2. 分析対象データ

2.1 分析対象プロジェクト

分析対象プロジェクトは、IPAで収集したエンタープライズ系ソフトウェアプロジェクト4,067件 [6]のうち、次の条件を満たすものを対象とする^{*4}。

- 1) システム稼働後の発生不具合数が報告されている。
- 2) 規模を示すFP (ファンクションポイント) の実測値 (5001FP実測値_調整前^{*5}) が報告されている。
- 3) 開発種別が改修・保守、再開発、拡張のいずれかである。

^{*3} 2018年6月の組織変更まで略称でSECと呼ばれていた。

^{*4} これらの条件を取り上げた理由については文献 [5] を参照されたい。

^{*5} 番号と名称はデータ白書 [6] で定義されたもの。

表 1 分析対象データの基本統計量

Table 1 Fundamental statistics of analyzed data

項目	改良開発		新規開発 (参考)	
	不具合数	FP規模	不具合数	FP規模
データ件数 (a)	155		305	
ゼロデータ件数 (b)	66	0	99	0
比率 (b/a)	0.43	0	0.32	0
平均	5.7	2.91	14.5	3.09
分散	157.3	0.214	4,086.1	0.213
最大	80	3.96	999	4.32
最小	0	1.18	0	1.93

(注) FP規模の値は、データ件数とゼロデータ件数以外は常用対数変換後の値

なお、本論文ではこれら3つの開発種別の和集合を改良開発と呼ぶことにする。

- 4) 開発5工程 (基本設計, 詳細設計, 製造, 結合テスト, 総合テスト (ベンダ確認)) をすべて実施しているプロジェクトである。

2.2 目的変数と説明変数

(1) 目的変数

目的変数は文献 [5] にならって、システム稼働後1ヶ月後の発生不具合現象数 (5267発生不具合現象数 (合計) 1ヶ月) を目的変数とする。以下では稼働後1ヶ月後の発生不具合現象数を単に不具合数と呼ぶ。

2.1の1)~4)を満たすプロジェクト数は155である。155件のデータの基本統計量を表1に示す。ただし、FP規模の基本統計量は、一般にFP規模が対数正規分布に従う [7] ため (常用) 対数変換後のものである。

なお、本論文では基本的に自然対数 (ln) を用いるが、図表など直観的なわかりやすさが重要と考えられる場合は常用対数 (log) も用いる。

表1から新規開発と改良開発では基本統計量に次のような違いが見られる。

- データ件数は新規開発の半分である。
- ゼロデータの占める割合は新規開発の32%に比べて43%と多い。
- 不具合数の平均は新規開発の14.5件に比べて5.7件と少ない。
- FP規模の平均 (対数変換前では中央値に相当する) は、新規開発の3.09 (対数変換前では中央値1,220FP) に比べて2.91 (対数変換前の中央値815FP) と小さい。

(2) 説明変数

ソフトウェアプロジェクトで扱う変数には、比尺度に従う量的変数と、名義尺度または順序尺度に従う質的変数がある。順序尺度に従う質的変数のうちから信頼性に影響を与える可能性のある変数として表2に示す51個を選んだ。

表 2 影響要因の候補 (順序尺度に従うもの)

Table 2 Candidates of effective qualitative factors subject to ordinal scale.

分類	変数	
プロジェクト全般 (P) (11)	111_新技術利用, 112_役割分担_責任所在, 113_達成目標_優先度_明確度合, 1011_定量的出荷品質基準_有無, 1013_第三者レビューの有無, 5241_品質保証体制_基本設計	
	114_作業スペース, 115_プロジェクト環境_騒音	
ツールの利用 (T) (11)	302_業務パッケージ, 403_類似プロジェクト, 404_プロジェクト管理ツール, 405_構成管理ツール, 406_設計支援ツール, 407_ドキュメント作成ツール, 408_デバッグ_テストツール, 409_CASE ツール, 411_コードジェネレータ, 412_開発方法論利用, 422_開発フレームワーク	
	501_要求仕様_明確さ	
ユーザー側 (U) (14)	要求仕様変更発生状況	5116_基本設計, 5117_詳細設計, 5118_製作, 5119_結合テスト, 5120_総合テスト (注 2)
	ユーザー担当者	502_要求仕様関与, 509_受け入れ試験関与, 503_システム経験, 504_業務経験, 507_設計内容理解度
	505_ユーザとの役割分担・責任所在_明確度合	
要求レベル (R) (8)	ユーザ承認有無	506_要求仕様, 508_設計
	512_信頼性, 513_使用性, 514_性能・効率性, 515_保守性, 516_移植性, 517_ランニングコスト要求, 518_セキュリティ	
開発担当者 (D) (7)	601_PM スキル	
	要員スキル	602_業務分野経験, 603_分析・設計経験, 604_言語・ツール利用経験, 605_開発プラットフォーム使用経験
	1010_テスト体制	スキルレベル, 要員数

(注 1) 番号はデータ白書 [6] で定義されたもの

(注 2) 正確には、「総合テスト (ベンダ確認)」であるが、本論文では単に「総合テスト」とする。

データ白書では類似の変数を 100 番台が同じものでグループ化しているので、これを参考にしながら、表 2 ではその変数の意味を考慮して変数を 5 つのグループに分類した。

名義尺度に従う質的変数のうち、代表的な変数として業種とアーキテクチャに属するものを取りあげる。この中で分析に耐えられるだけのデータ件数が揃っている 4 つの業種と 5 種類のアーキテクチャを表 3 に示す。業種は 201_業種_1 で示された値 (分類項目) を日本標準産業分類の大分類のレベルで分類し直したものの*6であり、アーキテクチャは 308_アーキテクチャ_1 に示された値 (分類項目) そのものである。具体的な変数名は、例えば「業種_製造業」のように表 3 の分類と変数を結び付けたものを用いる。

*6 データ白書 ([6]) の付録 A.3 参照

表 3 影響要因の候補 (名義尺度に従うもの)

Table 3 Candidates of effective qualitative factors subject to nominal scale.

分類	変数
業種 (I) (4)	製造業, 情報通信業, 卸売・小売業, 金融・保険業
アーキテクチャ (A) (5)	スタンドアロン, メインフレーム, 2 階層クライアントサーバ, 3 階層クライアントサーバ, インタネット・イントラネット

今回の分析では、1 個の質的変数に加えて、不具合数に大きな影響を与えていると思われる FP 規模をいわゆるコントロール変数として説明変数に加える。

分析結果の頑健性を高めるために、各説明変数は次の 3 つの条件を満たすものとする。

- 1) データ件数 (回答数) が 30 件以上ある。
- 2) 各レベルに属する回答数が 10 件以上ある。
- 3) 偏り率 ρ が $-0.7 \leq \rho \leq 0.7$ の範囲にある。

ただし、 ρ は N_1 を上位レベルの回答数、 N_2 を下位レベルの回答数としたとき

$$\rho = \frac{N_1 - N_2}{N_1 + N_2} \quad (1)$$

で定義したものである。 ρ のとり得る範囲は -1 から 1 であり、 $N_1 = N_2$ の場合は $\rho = 0$ となる。この条件は、回答数が上位レベルに偏る天井効果および下位レベルに偏る床効果を避けるために加えたものであり、全体の回答数に占める上位 (又は下位) のレベルの回答数の許容される割合は 15%~85%になる。

2.3 変数変換

(1) 量的変数

FP 規模は対数正規分布に従うので対数変換をする。対数は自然対数を用いる。不具合数は変数変換を行わずそのままの値を用いる。

(2) 順序尺度に従う質的変数

順序尺度に従う変数の多くは 3 つ以上のレベルの値をもつが、各レベル間が等間隔であるという保証はないため、厳密にはレベルを表す数をそのまま用いて回帰分析を行うことはできない。そこで今回の分析では 3 つ以上のレベルをとりうる変数は隣り合うレベル同士を合併して全体で 2 レベルにする (2 値化する)。このときレベルの若番の方を上位レベル、老番の方を下位レベルと呼ぶ。2 値化する分割点は複数個あるが、偏り率 ρ の絶対値が最も小さくなるものを分割点とする。本研究では下位レベルをベースになる群とみなすこととして、下位レベルに 0 を上位レベルに 1 の値を割り当てる。したがって、上位レベルの不具合数が少ない場合は回帰係数はマイナスとなる。

(3) 名義尺度に従う質的変数

名義尺度に従う変数に対しては、変数ごとに次のように

2 値化する。すなわち、対象とするプロジェクトが変数名に含まれる業種に該当する場合は、該当する業種名を、該当しない場合は「それ以外の業種」という値を割り当てる。例えば、変数「業種_製造業」の値は、「製造業」と「それ(製造業)以外の業種」の2つであり、前者を上位レベル、後者を下位レベルとする。業種欄が空欄(欠損値)の場合は、業種を分析する場合に限り、そのプロジェクトを分析対象から除外する。アーキテクチャの場合も同様である。

3. 分析方法

3.1 負の二項回帰モデル

目的変数がゼロを多く含むデータに対しては、一般化線形モデルに属するポアソン回帰モデルや負の二項回帰モデルを用いる方法が知られている(文献[8]の3.2, 3.3)が、今回の分析は文献[5]にならって負の二項回帰モデルを用いる。

一般化線形モデルでは、説明変数(群)の線形結合と目的変数の平均 μ をリンク関数で結合する。負の二項回帰モデルのリンク関数は対数関数である。

今回の分析では、説明変数が対数変換された FP 規模と、0 と 1 で 2 値化された質的変数 x の 2 つであることから次の式で不具合数の平均 μ を推定する。

$$\ln \mu = \beta_0 + \beta_1 \ln FP + \beta_2 x \quad (2)$$

計算は統計ツール R を用いて行う。

3.2 説明変数の選択基準

信頼性に影響を与える要因の選択は、回帰係数の傾きの p 値に基づいて行う。今回は質的変数が 2 値しかとらないので回帰係数の傾きは、上位レベルと下位レベルの 2 つの群のそれぞれの平均値の差と等しい。判断基準とする p 値の有意水準は、影響の可能性を幅広く確認するため、統計学で標準的に使われている 5% より高い 10% (両側検定) とする。

4. 分析結果

4.1 分析条件を満たさない変数

表 2 および表 3 であげた説明変数の候補のうち、次の 16 個の変数が 2.2 の (2) で述べた説明変数が満たすべき 3 つの条件のうち、少なくともひとつの条件を満たさないことがわかった。

- プロジェクト全般：111_新技術の導入*、1013_第三者レビューの有無、115_プロジェクト環境_騒音、計画の評価(120_コスト、121_品質*、122_工期)。
- ツールの利用：409_CASE ツール。
- ユーザ側：5119_要求仕様変更発生状況_結合テスト*、5120_要求仕様変更発生状況_総合テスト(ベンダ確認)*、506_要求仕様_ユーザ承認有無、508_設計_ユー

ザ承認有無*。

- 業種：情報通信業、卸売・小売り業。
- アーキテクチャ：スタンド・アロン、メインフレーム、2 階層クライアントサーバ。

*印をつけた変数は新規開発では分析条件をすべて満たしていた変数である。逆に新規開発では分析条件を満たさなかったが、改良開発では条件を満たした変数は 405_構成管理ツールの利用である。

条件を満たさなかった 16 個の変数は分析の対象外とし、以降は 44 個の変数を分析対象とする。

4.2 信頼性への影響要因

信頼性への影響要因として選択された変数、すなわち 2.2 の (2) で述べたデータ件数に関する 1)~3) の条件、および 3.2 の説明変数の選択基準をすべて満たす 6 個の質的変数を表 4 に示す。なお、2.2 の (2) で述べたように、回帰分析の説明変数は FP 規模とひとつの質的変数であることから、表 4 に示す質的変数の偏回帰係数は、FP 規模の影響を除去した後 (FP 規模が一定とした場合) の目的変数への影響の大きさを示していることに留意してほしい。

表 4 から次のことがわかる。

- プロジェクト全般では、影響要因となるものはない。
- ツールを利用することにより信頼性が高くなるものは、408_デバッグ_テストツールである。
- ユーザ側では、502_ユーザ担当者が要求仕様に関与している場合、503_ユーザ担当者のシステム経験が豊富な場合、507_ユーザ担当者の設計内容理解度が高い場合が、いずれも信頼性が高い。
- 要求レベルで影響要因となるものはない。
- 開発担当者に関しては、1010_テスト体制でテスト要員数が少ない方が信頼性は高い。
- 業種では、金融・保険業が他の業種に比べて信頼性が高い。
- アーキテクチャでは、他のアーキテクチャと信頼性に差があるものはない。

5. 考察

5.1 不具合数比率

「不具合数比率」は、ある FP 規模における上位レベルおよび下位レベルそれぞれの不具合数の平均をそれぞれ μ_i^+ および μ_i^- とした場合の μ_i^+ と μ_i^- の比であり、

$$\frac{\mu_i^+}{\mu_i^-} = e^{b_2} \quad (3)$$

で表すことができる [5]。ただし、 b_2 は式 (2) における β_2 の推定値である。

表 4 に影響要因の候補の不具合数比率を示す。表 4 によると、1010_テスト体制(要員数)を除くすべての不具合数比率が 0.6 以下である。

表 4 影響要因とみなされた質的変数
Table 4 Qualitative variables identified as effective factors.

分類	変数名		内容		データ件数			質的変数		不具合数比率
			上位レベル	下位レベル	総数	上位レベル	下位レベル	偏回帰係数	p 値 (%)	
T	408_デバッグ_テストツール利用		a : 有り	b : 無し	91	28	63	-0.70	8.3	0.49
U	ユーザ担当者	502_要求仕様関与	a : 十分に関与	b : 概ね関与 + c : 関与が不十分 + d : 未関与	83	24	59	-1.00	1.2	0.37
		503_システム経験	a : 十分に経験	b : 概ね経験 + c : 経験が不十分 + d : 未経験	76	29	47	-1.32	0.1	0.27
		507_設計内容理解度	a : 十分に理解	b : 概ね理解 + c : 理解が不十分 + d : 全く理解していない	75	33	42	-0.67	7.4	0.51
D	1010_テスト体制 (要員数)		a : スキル要員ともに十分 + c : スキルは不足, 員数は十分	b : スキルは十分, 員数は不足 + d : スキル, 員数ともに不足	48	30	18	0.84	5.4	2.33
I	業種_金融・保険業		金融・保険業	左記以外の業種	155	67	88	-0.51	9.2	0.60

5.2 外れ値の影響と多重共線性

改良開発プロジェクトでの不具合数は大きい順に 88 件, 70 件, 54 件, 49 件と続き, 外れ値かどうかを検討すべき他と離れた大きな値のものはなかった。

また, 6 つのの影響要因のうち, χ^2 検定で独立ではないと判定されたいずれもユーザ側に属する 3 つの要因について多重共線性を調べたところ (データ件数は 75 件), VIF (Variance Inflation Factor) は 1.0 から 1.5 の範囲にあり, いずれも多重共線性はないという目安となる $VIF < 10$ の基準を満たしていることから, 6 つの影響要因に冗長性はないと言える。

5.3 常識と異なる要因の詳細検討

表 4 のうち, 1010_テスト体制 (要員数) が要因であるという結果, すなわち, 十分なテスト要員数が確保された場合の方がそうでない場合よりも信頼性が低下するという結果は一般的な常識と反する。このような結果が得られた理由を明らかにするために, 業種の影響と量的変数の影響を調べた。調査対象は 1010_テスト体制 (要員数) だけでなく, 表 4 で示したすべての影響要因とした。

(1) 業種の影響

業種_金融・保険業を除く 5 つの要因に対して, 有意水準 5% で業種_金融・保険業との独立性の検定 (χ^2 検定) を行ったところ, 1010_テスト体制 (要員数) と金融・保険業の組合せだけが有意 (p 値は 3.7%), すなわち独立でないことがわかった。2 つの要因の ϕ 係数は -0.34^{*7} , すなわちテスト体制で要員数が十分なプロジェクトは, 一般に信

頼性の高い金融・保険業には有意に少ないということがわかった。

(2) 量的変数の影響

線形回帰分析の場合, 変数 x_1 が目的変数に影響を与えていない場合でも, それが目的変数に影響を与えている変数 x_2 と相関がある場合には, x_1 の単回帰係数が見かけ上有意となることがある。このような場合, x_1 と x_2 を説明変数とした重回帰分析を行うと, x_1 の偏回帰係数の絶対値は小さくなって x_1 が見かけ上の影響要因であることがわかる。

負の二項回帰分析の場合にも同様な可能性が考えられるので, 表 4 に示した変数 (x_1 に相当) に対して, それと不具合数に影響を与える量的変数 (x_2 に相当) に相関がある場合, それら 2 つの変数にコントロール変数としての FP を加えた重回帰分析を行って表 4 に示した変数の回帰係数の p 値の変化を調べた。

追加した量的変数は, データ件数が十分でかつ不具合数への影響要因 (回帰係数が有意) である, 検出バグ現象数 (結合テスト) と同 (総合テスト) の 2 つを選んだ。

(3) 重回帰分析

表 4 に示した 6 つの要因のうち, 1010_テスト体制 (要員数) を除く 5 つの要因に対しては, 上記 2 つの量的変数のすべての組合せ (3 通り: 量的変数ひとつの場合が 2 通り, 2 つの量的変数を共に加えた場合が 1 通り) を追加した重回帰分析を, 1010_テスト体制 (要員数) に対しては, 2 つの量的変数に業種_金融・保険業を加えた 3 つの変数のすべての組合せ (7 通り) を加えた重回帰分析を行った。

分析結果のうち意味のある組合せを表 5 に示す。表 5 には, 偏回帰係数のゆらぎを考慮して p 値が 10% から 15% のものも検討対象に加えた。

*7 ϕ 係数を 2×2 の分割表におけるクラメールの連関係数 (χ^2 値) とした場合は正の値しかとらないが, 四分点相関係数 (ピアソンの相関係数と等しい) を ϕ 係数とした場合は負の値もとれる。本論文では後者の定義に従う。

1010_テスト体制(要員数)についてみると、業種_金融・保険業だけを加えた場合の回帰係数とその p 値は、1010_テスト体制(要員数)が 0.78 と 9.8%、業種_金融・保険業が -0.69 と 10.8% となった。テスト体制(要員数)の回帰係数は単独の場合の 0.84 より小さくなったが、金融・保険業だけによる影響は小さい。金融・保険業と検出バグ現象数(総合テスト)を加えるとテスト体制(要員数)の回帰係数は 0.26 で p 値は 50% を超えている。この結果から 1010_テスト体制(要員数)は要因であるとは言い切れない。

デバッグ_テストツールの利用は検出バグ現象数(総合テスト)の影響を強く受けているが、偏回帰係数は -0.68 でその p 値は有意水準と定めた 10% よりわずかに大きい 11.6% であり、これだけからこの変数が要因ではないとは断定できない。

以下では 1010_テスト体制(要員数)を除く 5 つの要因について議論を進める。

5.4 新規開発プロジェクトとの違い

5.4.1 信頼性への影響要因の違い

新規開発と改良開発の影響要因の比較を表 6 に示す。表 6 には、各影響要因のもう一方の開発種別での p 値が 10% 台のものも参考のために記載している(例えば改良開発の要因である 503_ユーザ担当者_システム経験の新規開発での偏回帰係数と p 値)。

表 6 から改良開発プロジェクトでの影響要因は新規開発での影響要因の概ねサブセットであることがわかる。

- プロジェクト全般では、113_達成目標_優先度_明確度が要因とはならなかった。
- ツール利用では 408_デバッグ_テストツールが新規開発と同様に要因となった。
- ユーザ側に関しては、ユーザ担当者の 502_要求仕様への関与、および 507_設計内容理解度が新規開発と同様に要因となった。
- 新規開発では p 値が 12.8% のため要因とは選定されなかった 503_ユーザ担当者_システム経験が改良開発で要因となったが、501_要求仕様_明確さ、504_ユーザ担当者_業務経験、5116_および 5117_上流工程における要求仕様変更発生の抑制は要因とはならなかった。
- 要求レベルの高さは新規開発とは異なり、影響要因とはならなかった。ただし、515_要求レベル_保守性の偏回帰係数は -0.54、p 値は 15.3% であり、改良開発プロジェクトが要求レベルの影響を全く受けないわけではない。
- 開発担当者に関しては、1010_テスト体制(スキルレベル)が要因とはならなかった。

改良開発では既に開発されたソフトウェア(母体)があることが新規開発と異なる。開発済の母体があるということは、ソフトウェアの開発目的構造・開発方針などは確立

しているということでもある。それが新規開発との影響要因の差の一因と考えられる。

例えば、改良開発では達成目標_優先度は、たとえその明確度が低いという場合でも実際には既に確立されていてそれに倣えばよいという可能性がある。また、改良開発における要求仕様は、母体のソフトウェア構造を揺るがすようなものではなく、ユーザ担当者が要求仕様をしっかり関与していれば信頼性は要求仕様のゆらぎには大きく影響されないことが考えられる。的外れな要求仕様の作成を減らすという点で有効なユーザ担当者の業務経験の有無は、改良開発の場合は母体があるので重要度は低い可能性がある。さらに移植性や保守性に対する高い要求レベルも、母体のソフトウェア構造が確立されていれば、その影響は少ない可能性がある。テストでは新規開発と改良開発では大きく異なり、改良開発のときのテストは新規開発のときのテストケースやテスト担当者など多くを新規開発時のテストを踏襲することが多く、レベルが低い担当者でも改良開発のときはそれほど問題にならずにテストをすることができる。

以上のことから、改良開発で信頼性の向上を図るためには、ユーザとのコミュニケーションを密にすること、すなわちユーザ担当者が要求仕様作成に深く関与してユーザ要求に即した仕様を柔軟に作成すること、開発にあたってはユーザ担当者も設計内容を理解し、豊富なシステム経験から開発を支援することが重要である。また開発にあたっては、デバッグ_テストツールを利用が効果がある。

5.4.2 開発プロセスの変化と信頼性への影響

表 1 からわかるように、改良開発は新規開発に比べて不具合数が少ないが FP 規模も小さい。不具合数を目的変数、新規開発を 0 とし、改良開発を 1 とした開発種別、およびコントロール変数としての FP 規模の 2 つを説明変数とする回帰分析を行うと、開発種別の偏回帰係数が -0.37 (不具合数比率は 0.69)、p 値が 4.1% となり、FP 規模の影響を除いても改良開発プロジェクトは新規開発プロジェクトよりも不具合数が少ない(信頼性が高い)ことがわかる。

その理由を調べるために、質的変数から見た新規開発プロジェクトと改良開発プロジェクトの開発プロセスの違いを調べた。順序尺度に従う質的変数は、5241_品質保証体制_基本設計*8を除いて上位レベルがプロセスとしてのレベルが高いと考えられる。従って下位レベルに対する上位レベルの回答数の比率が高くなった変数はその変数に関するプロセスが「進化(改善)」したと言える。

新規開発と改良開発それぞれにおける質的変数の上位レベルと下位レベルの回答数の比率に統計的有意差の有無を調べた。すなわち、1 行目を改良開発、2 行目を新規開発、1 列目を上位レベルと 2 列目を下位レベルとする 2 × 2 のクロス集計表を作成してそれに対する χ^2 検定(独立性

*8 a: プロジェクトメンバが実施, b: 品質保証の専門スタッフが実施である。

表 5 業種と量的変数を含む重回帰分析の結果

Table 5 Multivariate regression analysis results including the type of industry and quantitative variable.

要因名	データ件数			偏回帰係数 (p 値 (%))			(重) 相関係数 (p 値 (%))
	総数	上位レベル	下位レベル	変数 (要因)	金融・保険業	検出バグ現象数 (総合テスト)	
1010_テスト体制 (要員数)	48	30	18	0.78 (9.8)	-0.69 (10.8)	—	-0.34 (1.6)
	42	26	16	0.26 (50 <)	-0.70 (14.0)	0.42 (3.9)	0.47 (0.1)
408_デバッグ_テストツール利用	69	26	43	-0.68 (11.6)	—	0.59 (0.0)	-0.20 (9.7)

(注) 説明変数には上記の変数以外にコントロール変数として FP 規模を含む

表 6 新規開発プロジェクトと改良開発プロジェクトの要因の比較

Table 6 Comparison of factors between newly developed and modified software projects

分類	変数	新規開発			改良開発			
		データ件数	偏回帰係数	p 値 (%)	データ件数	偏回帰係数	p 値 (%)	
P	113_達成目標_優先度明確度合い	160	-0.50	9.4	—			
	121_計画の評価 (品質)	249	-1.02	0.1	(偏り率が 0.71)			
T	408_デバッグ_テストツールの利用	139	-0.83	1.2	91	-0.70	8.3	
U	501_要求仕様の明確さ	193	-0.60	1.2	—			
	要求仕様変更発生状況	5116_基本設計	46	-1.17	3.2	—		
		5117_詳細設計	45	-1.38	0.9	—		
	ユーザ担当者	5119_結合テスト	36	-1.51	1.3	(下位レベルの回答数が 7 件)		
		5120_総合テスト	42	-0.89	8.9	(下位レベルの回答数が 9 件)		
	R	要求レベル	502_要求仕様関与	167	-0.64	0.8	83	-1.00
503_システム経験			97	-0.57	12.8	76	-1.32	0.1
504_業務経験			84	-0.72	6.1	—		
507_設計内容理解度			89	-0.97	3.0	75	-0.67	7.4
D	1010_テスト体制 (スキルレベル)	514_移植性	122	-0.74	2.2	—		
		515_保守性	88	-1.13	0.8	76	-0.54	15.3
I	業種_金融・保険業	305	-0.66	0.3	155	-0.51	9.2	

(凡例) (): 分析対象外とした理由, — : p 値が 20%を超えたもの

の検定)を行った。φ 係数が正の場合は改良開発の上位レベルの回答比率が新規開発よりも高くなり、負の場合は新規開発よりも低くなる。

有意水準 5%で有意となった組合せの φ 係数とその P 値を表 7 に示す。表 7 から次のことがわかる。

- プロジェクト全般では概ね変化は見られなかったが、111_新技術の導入が減少し、112_役割分担_責任所在の明確化は向上している。
- ツール利用という点では、403_類似プロジェクトの利用は増加しているが、404_プロジェクト管理ツール、406_設計支援ツール、408_デバッグ_テストツールの利用率がいずれも低下している。
- ユーザ側および開発担当者に属する変数で多くの組合せで有意となったが、いずれも開発プロセスは進化(改善)している。
- 要求レベルに属する変数で有意となったものはない(要求レベルは新規開発でも改良開発でも変化はない)。

- 業種では、卸売・小売業のプロジェクトが減少し、金融・保険業のプロジェクトが増加している。

全体としてみると、ツールの利用を除いて改良プロジェクトは新規開発プロジェクトよりも開発プロセスは進化していることがわかる。このことが改良開発プロジェクトの信頼性が新規開発プロジェクトの信頼性よりも高い理由のひとつと考えられる。

ツール利用では、新規開発でも改良開発でもデバッグ_テストツールの利用が信頼性の向上に寄与することから、特段の事情がない限り、改良開発でもデバッグ_テストツールを少なくとも新規開発並みに利用すべきであると言える。

6. おわりに

今回、IPA の収集したデータをもとに改良開発プロジェクトにおける信頼性への影響要因を分析した。分析方法は新規開発プロジェクトを対象とした場合と同様の方法をとって結果を比較した。改良開発プロジェクトにおける信

表 7 新規開発プロジェクトと改良開発プロジェクトで質的変数に対する回答数の比率が異なるもの

Table 7 Qualitative variables that the ratio of the number of answers between the upper and lower levels is different between new and modified development projects.

分類	変数	内容		データ件数(回答数)とその比率						χ^2 検定結果				
		上位レベル	下位レベル	新規開発			改良開発			ϕ 係数	p 値 (%)			
				上位レベル	下位レベル	比率(注)	上位レベル	下位レベル	比率(注)					
P	111_新技術の導入	a	b	46	135	0.25	17	111	0.13	-0.15	0.9			
	112_役割分担_責任所在	a	b+c+d	64	98	0.40	54	48	0.53	0.13	3.3			
T	403_類似プロジェクト	a	b	56	58	0.49	65	26	0.71	0.23	0.1			
	404_プロジェクト管理ツール			125	34	0.79	70	35	0.67	-0.13	3.1			
	406_設計支援ツール			62	77	0.45	20	75	0.21	-0.24	0.0			
	408_デバッグ_テストツール			67	72	0.48	28	63	0.31	-0.17	0.9			
U	要求仕様 変更発生 状況	a	b+c+d	5117_詳細設計	16	29	0.36	24	17	0.59	0.23	3.3		
				5119_結合テスト	19	17	0.53	30	7	0.81	0.30	1.0		
				5120_総合テスト	24	18	0.57	32	9	0.78	0.22	4.2		
	ユーザ 担当者	a	b+c+d	502_要求仕様関与	a+b	c+d	100	67	0.60	63	20	0.76	0.16	1.2
				503_システム経験	a	b+c+d	17	80	0.18	29	47	0.38	0.23	0.2
					a+b	c+d	59	38	0.61	59	17	0.78	0.18	1.8
				507_設計内容理解度	a	b+c+d	19	70	0.21	33	42	0.44	0.24	0.2
	a+b	c+d	69		20	0.78	69	6	0.92	0.20	1.1			
505_ユーザとの役割分担_責任所在_明確度合	a+b	c+d	84	16	0.84	77	5	0.94	0.15	3.7				
D	要員 スキル	a	b+c+d	602_業務分野経験	a+b	c+d	97	50	0.66	89	19	0.82	0.18	0.4
				603_分析_設計経験	a	b+c+d	25	103	0.20	36	58	0.38	0.21	0.2
				604_言語_ツール利用経験	a	b+c+d	27	103	0.21	34	61	0.36	0.17	1.2
				605_開発プラットフォーム 使用経験	a	b+c+d	31	98	0.24	37	56	0.40	0.17	1.2
I	業種_卸売_小売業	卸売_小売業	左記以外	39	266	0.13	10	145	0.06	-0.10	3.7			
	業種_金融_保険業	金融_保険業	左記以外	90	215	0.30	67	88	0.43	0.14	0.3			

(注) 全回答数に対する上位レベルの回答数の比率

信頼性への影響要因は、概ね新規開発プロジェクトで抽出されたもののサブセットであり、信頼性の向上にはユーザ側の協力が重要であることが改めて浮き彫りになった。また、改良開発プロジェクトの方が、新規開発プロジェクトに比べてプロジェクト規模が小さいことを考慮にいれても信頼性が高いこと、その理由のひとつとして改良開発プロジェクトが新規開発プロジェクトよりも開発プロセスが改善していることも明らかにした。ただし、信頼性の向上に有効なデバッグ_テストツールの利用が改良開発プロジェクトで低下しているので、信頼性を重視するならばその利用を新規開発のときのように進めるのがよいと思われる。

謝辞 本研究は東海大学と IPA/社会基盤センターが連携して実施したものである。IPA/社会基盤センターの山下リーダー並びに五味研究員のご協力を深く感謝致します。

参考文献

[1] Zhang, X. and Pham, H.: An Analysis of Factors Affecting Software Reliability, *J. Sys. Softw.*, Vol.50, No.1,

pp.43-56 (2000).
 [2] Zhu, M., Zhang, X. and Pham, H.: A Comparison Analysis of Environmental Factors Affecting Software Reliability, *J. Sys. Softw.*, Vol.109, pp.150-160 (2015).
 [3] Zhu, M. and Pham, H.: Environmental Factors Analysis and Comparison Affecting Software Reliability in Development of Multi-Release Software, *J. Syst. Softw.*, Vol.132 (C), pp.72-84 (2017).
 [4] Gorla, N. and Lin, S-C.: Determinants of Software Quality: A survey of Information Systems Project Managers, *Informat. Softw. Technol.*, Vol.52, No.6, pp.602-610 (2010).
 [5] 古山恒夫: エンタープライズ系ソフトウェアの信頼性に影響を与える質的要因の分析, 情報処理学会論文誌, Vol. 60, No. 11, pp. 2048-2062 (2019).
 [6] 独立行政法人情報処理推進機構 (IPA) ソフトウェアエンジニアリングセンター (SEC) 監修: ソフトウェア開発データ白書 2016-2017 (2016).
 [7] 古山恒夫: ソフトウェアプロジェクトデータの量的変数に関する分析の一指針と分析事例, SEC journal, 第7巻, 第3号 (通巻 26号), pp. 105-111 (2011).
 [8] Cameron, A. C. and Trivedi, P. K.: *Regression Analysis of Count Data*, 2nd ed., Cambridge Uni. Press, p.566 (2013).