

ソフトウェア開発における 品質管理の理論と実践

- 欠陥件数の持つ代数的性質とその品質管理への応用

2016/10/12

(株)プロセス分析ラボ

小室 睦

● 株式会社プロセス分析ラボ

■ 企業理念

- ◆ ソフトウェアの開発・管理のプロセスを科学的に分析し、自律的な改善をサポートしていくことで、ソフトウェア開発者が創造性を発揮し、より価値の高いシステムをお客様にとどけられるようにいたします。

■ 主な事業内容

- ◆ ソフトウェア開発プロセスの現状評価(CMMI®アプレイザル)、改善コンサルティング
- ◆ お客様の状況にあった改善状況・改善効果の見える化のサポート(指標定義、分析)
- ◆ データ分析など科学的手法に基づいた効果的な改善活動のコンサルテーション

● 発表者紹介: 小室 睦

- ◆ 博士(情報科学)
- ◆ CMMI 高成熟度リードアプレイザ

■ 主な論文・著作

- ◆ 「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) **日経品質管理文献賞**
- ◆ 「開発現場の実態に基づいたピアレビュー手法改善と改善効果の定量的分析」, *SEC Journal*, Vol.1, No.4, pp.6-15 (2005) **SECジャーナル創刊記念論文最優秀賞**
- ◆ 「ソフトウェア開発における予測モデル構築手法の提案」 **SPI Japan 2011 PC委員長賞**
- ◆ “Experiences of Applying SPC Techniques to Software Development Processes,” in *Proc. of ICSE'06*, pp.577-584 (2006).



- 本プレゼンテーションでは既に発表済みの一連の論文・著作の内容に基づき、主に発表者の経験を報告する
- 小室睦、薦田憲久:「ピアレビューデータに基づく品質予測モデル」、電子情報通信学会論文誌D, Vol. J94-D, No. 2, pp.439-449 (2011)
- 野中誠、小池利和、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)の第5章
- 小室睦:「ソフトウェア開発における欠陥データ分布の理論とその品質管理への応用」情報学会デジタルプラクティス(投稿中)
- 加藤 拓海:「途中工程データが欠落した状況における影響評価モデルの構築」、SQiPシンポジウム(2016)

こんなことはありませんか？

- 改善効果の定量化
 - トップから改善活動の効果を問われても、明確に答えられない
 - 実施する改善施策の選択が勘にたよっている。実施効果も主観的にしかわからない。
- プロジェクト管理（品質管理）
 - 設計レビューが十分かどうか不安もあったが、計画日程が気になり、「まあ大丈夫だろう」と次工程にすすめた。結局バグが多発して納期を守れなかった。
 - 計画通りにプロジェクトを進めたが、上流工程での品質作り込みが悪くトラブルだった。よく考えたら計画の下敷きにした前プロジェクトでも同じ間違いをしていた。
- 提案する解決策
 - 上流工程での品質向上施策の効果を定量化
 - **品質向上効果の説明モデル、品質予測モデル**

- プロジェクト管理では上流工程が重要（これは常識）
- しかし、単に「重要」というだけでは、具体的なアクションにつながりにくい

- 定量化の必要性
 - （例）上流でレビューを徹底すれば下流工程（テスト）が楽になる
 - 問題点1: どこまで「徹底」すればいいのかわからない
 - 問題点2: どの程度「楽」になるのかわからない
 - 結果: スケジュールの圧力に負けてレビューが不十分なまま次工程に進むことに

- 単純に考えると
 - レビューでも不具合を抽出する
 - テストでも不具合を抽出する
 - 両者に何かの関係(相関)があるのでは？
- 実際にやってみるとうまくいかない
 - (理由1)正の相関と負の相関が混在
 - ◆ 正の相関となる要因
 - 開発の難易度が高い、もしくは同種のシステムの開発経験が少ない
 - ◆ 負の相関となる要因
 - 上流で通常以上に徹底したレビューを実施して、それまで下流で抽出していた不具合まで前倒しで抽出した
 - (理由2)レビューやテストのやり方だけで品質は決まらない
 - ◆ 他の要因を考慮して、「雑音」を消す必要

● 製品品質とプロセス品質

- 最終的には製品品質が問題となるが、製品品質は開発途中（特に上流工程）では適切に把握できない
- プロセス改善活動の前提

高いプロセス品質 ⇒ 高い製品品質

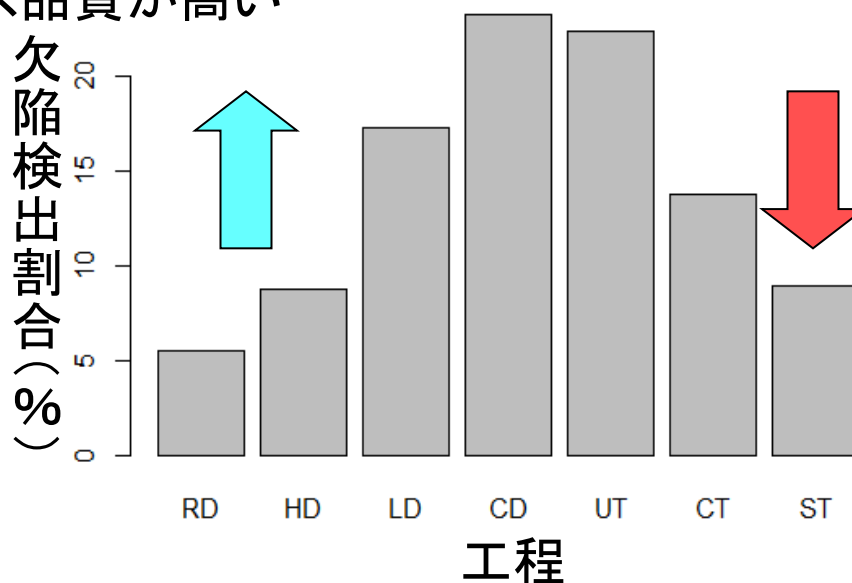
● 製品品質への影響

- 開発時の製品品質は伝統的にシステムテスト(ST)工程での欠陥件数、もしくは欠陥密度で測定されている
- 基本的仮説

プロセス品質の向上 ⇒ ST工程の欠陥件数減少

- ◆ この仮説を定量化した形で確かめたい

- 製品品質の時と同様、(上流工程でも)欠陥件数や密度は測定可能。ただし、意味はやや不明確
- 解決のアイデア: 工程別の欠陥検出件数を総欠陥数で割って各工程での検出割合を出す
 - 1つの確率分布だと思う
- グラフの形状がプロセス品質をあらわす
 - ピークが前に来ているとプロセス品質が高い
 - 上流で割合を高めれば下流で下がる
 - ◆ 定性的には当たり前、経験とも一致



- ウォータフォール型 (V字) のライフサイクルモデルを仮定
 - ◆ RD: 要件定義
 - ◆ HD: 高位設計
 - ◆ LD: 下位設計
 - ◆ CD: 実装
 - ◆ UT: 単体テスト
 - ◆ CT: 結合テスト
 - ◆ ST: システムテスト
- RD<HD<...<CT<STという順序がついているとみなす
- 工程 p で抽出された欠陥件数を $f(p)$ と書く

● 工程別の不具合件数

- 計画時に品質計画として目標設定する
- プロジェクト実施中の管理に活用
- グラフの形状から品質作り込み具合がわかる

● 工程別の不具合検出割合

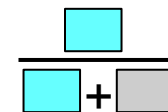
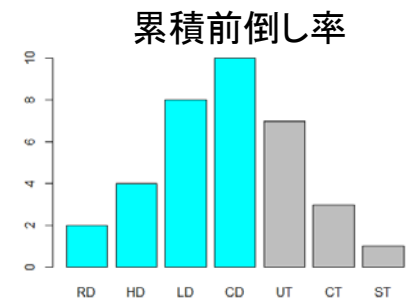
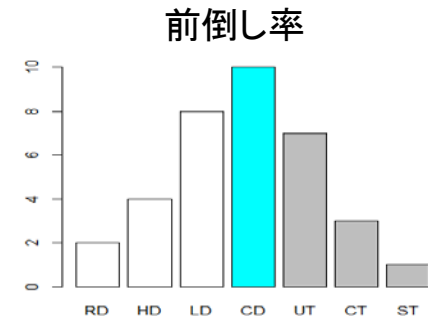
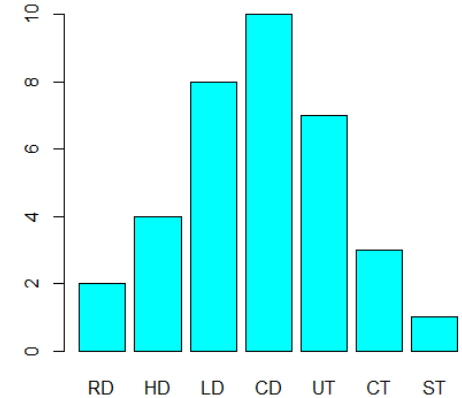
- 組織レベルでの改善状況の把握
 - ◆ (現状の)標準的な工程別検出割合を作成しておく改善やプロジェクト管理に有用
- プロジェクト間の比較に利用可
- 完了プロジェクトでないと描けない

● 工程別に表示した開発規模あたり不具合件数

- プロジェクト間の比較もでき、実施中のプロジェクトでも使用可

● 単一の数値(指標)で状況を把握したいとき

- 前倒し率、累積前倒し率などが有用
- これらも組織レベルの改善に利用可能



前倒し率と残存欠陥率の関係

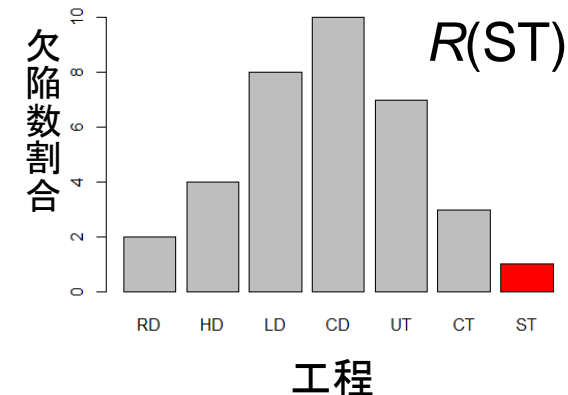
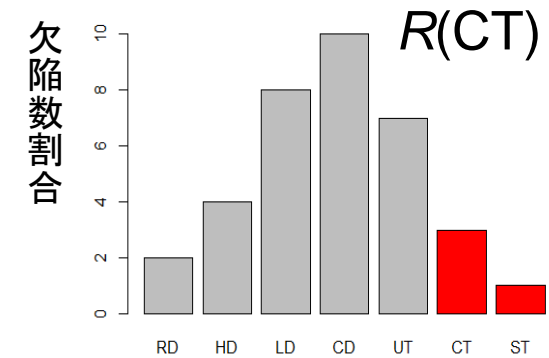
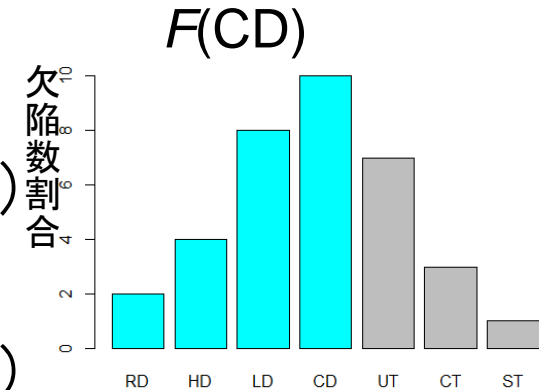
- 工程pでの累積前倒し率 $F(p)$
 - (p以前に摘出された欠陥数) / (総欠陥数)
- 工程pでの残存欠陥率 $R(p) = 1 - F(p')$
 - (p以後に摘出された欠陥数) / (総欠陥数)
 - p'はpの直前工程

● 仮説0

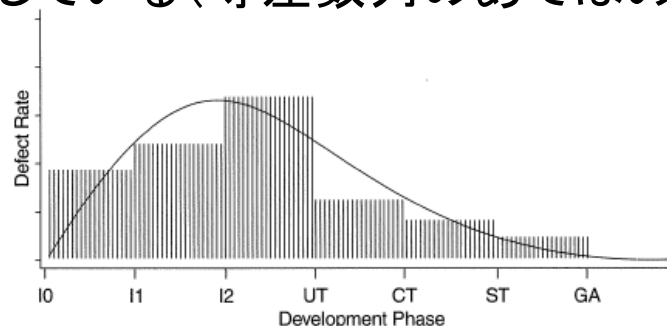
- 工程別欠陥件数分布はべき乗減衰則を満たす:

$$R(q) = R(p)^k \quad (q > p)$$

- ◆ 残存欠陥はべき乗で減っていく
- ◆ 係数 k は工程だけで決まる定数



- ソフトウェア工学への信頼性工学手法の適用例
 - Putnam: 工数の時系列分布をレイリー分布でモデル化
 - Gaffney, Kan: 工程別欠陥摘出数へのレイリー分布の適用
- 問題点・疑問点
 - 手持ちの欠陥件数データで試すと、レイリー分布から、ほど遠い場合も多い。
 - ◆ レイリー分布はワイブル分布で形状パラメータが一定値 2 の場合であるが、形状も変化しているように見える
 - 欠陥件数データでは時間パラメータの精度が低く、工程別にまとめた離散データになってしまう。
 - ◆ Kanの本では工程という順序尺度を無理やり間隔尺度として扱って適用している(等差数列のあてはめ)



出典: Kan, "Metrics and Models in Software Quality Engineering," Addison-Wesley, 2003.

● ワイブル分布

- 累積確率が次式で与えられる分布

$$F_{\alpha, \beta}(t) = 1 - \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]$$

◆ α : 形状パラメータ、 β : 尺度パラメータ

- 最弱リンクモデルの考察から導入された分布
- システムの弱い部分から欠陥が発生すると考えると、欠陥件数の分布としてワイブル分布を期待するのは自然
- 形状パラメータを一定値 α_0 としたワイブル分布は次のべき乗減衰則を満たす。

● 定義【べき乗減衰則】

- パラメータ β を含む確率分布族 $\{f_\beta(t)\}_\beta$ がべき乗減衰則を満たす
⇔

任意の時刻 t_1, t_2 に対して、 β に依存せず、 t_1, t_2 だけで決まる定数 c が存在して、信頼度関数 $R(t) = 1 - F(t)$ が次の等式を満たす

$$R(t_1) = R(t_2)^c$$

- $\log R(ST) = k \log R(p)$,
pはST以前の工程(LD, CD, UTなど)
- $\log f(ST) - \log(\text{total.defect}) = k \log R(p)$
 - total.defect は開発工程全体にわたる総欠陥摘出数
- $\log f(ST) = k \log (1-F(p')) + \log(\text{total.defect})$
 - $F(p')$ は工程p以前でのプロセスの品質をあらわしており、制御可能と考える
 - 右辺により左辺が計算されると読みたいが、total.defectは $f(ST)$ も含んでいるので、このままでは意味がない。
- 経験的に total.defectは開発規模sizeで80-90%決定される
 - 単回帰分析で検証
- p' が上流工程のとき、 $F(p')$ の値は小さいので、 $\log(1-F(p')) = -F(p')$ と置換え(近似)可能
- 仮説1(効果説明モデル)
 - $\log f(ST) = -k F(p') + A * \log(\text{size}) + (\text{他の要因からくる項})$
 - 他の要因としては、難易度、経験度などが考えられる

- 効果説明モデル

- $\log f(\text{ST}) = -k F(p') + A^* \log(\text{size}) + (\text{他の要因からくる項})$

- 課題: 効果説明モデルはプロセス品質と製品品質の間の定量的関係を表現しており有用であるが、予測能力はない

- 理由: $F(p') = \sum_{q < p} f(q) / \text{total.defect}$ の右辺にtotal.defectが含まれているため、開発完了までまたなければ実際の値が確定しない

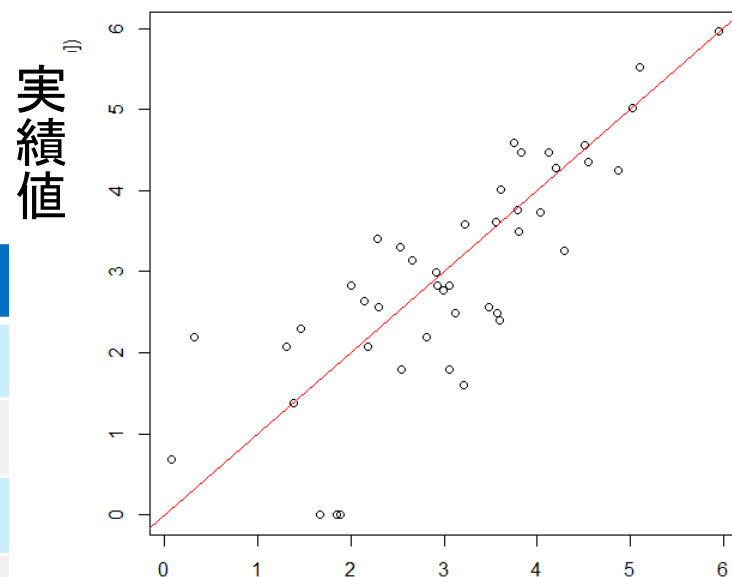
- 解決策: total.defct をsizeの定数倍で置換えて近似

- 仮説2(品質予測モデル)

- $\log f(\text{ST}) = -k' \sum_{q \leq p} f(q) / \text{size} + A' * \log(\text{size}) + (\text{他の要因からくる項})'$

- 重回帰分析(ポアソン分布を仮定した一般化線形回帰)で検証
 - ただし、過分散が観察されたため、quasipoissonを指定(R言語, glmパッケージを使用)
- 対象工程 p= HD, LD, CD, or UT
- 47件のプロジェクトデータ
- 目的変数 ST工程の欠陥摘出件数 (log をリンク関数として指定)
- 説明変数
 - 開発規模(の対数)
 - 累積前倒し率 F(p)
 - 業務種別を表すダミー変数
 - 改造規模(の対数): 難易度をあらわす
- 擬似決定係数、p値など

モデルの計算結果と実績値(LD)



モデル対象工程	擬似R ²	F(p)の係数	F(p)のP値
HD	86.1%	-3.48	2.2%
LD	87.8%	-2.96	0.2%
CD	85.9%	-2.46	3.0%
UT	91.9%	-5.57	0.1%未満

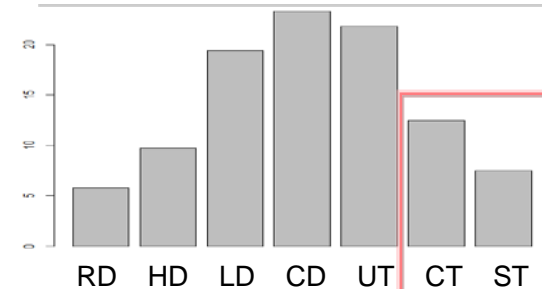
計算結果

- (1) 過去データから典型的な実績データ(現状の実力値)を表示
- (2) 改善施策とその効果の設定
 - 過去実績(標準分布、ベースライン)との比較、欠陥の原因分析などから改善すべき工程を特定
 - **実施可能な施策とその想定効果(実施工程での欠陥検出数アップ)を特定**
(実施例) 弱点分析(欠陥データの原因分析)から詳細設計書レビューを改善
 - ◆ 知見者のレビュー参加
 - ◆ レビュー速度、指摘密度の管理図を用いた制御
- (3) モデルを用いて、テスト工程への影響を定量的に評価

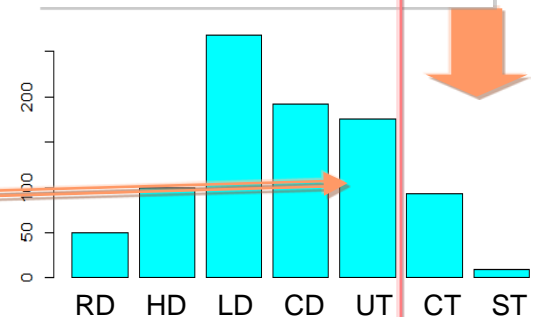
実現可能かつ効果のある品質計画を立案(具体的な施策の検討)

実施結果: CT工程以降の不具合密度が過去実績の半分未満に減少

部署の標準分布(イメージ)

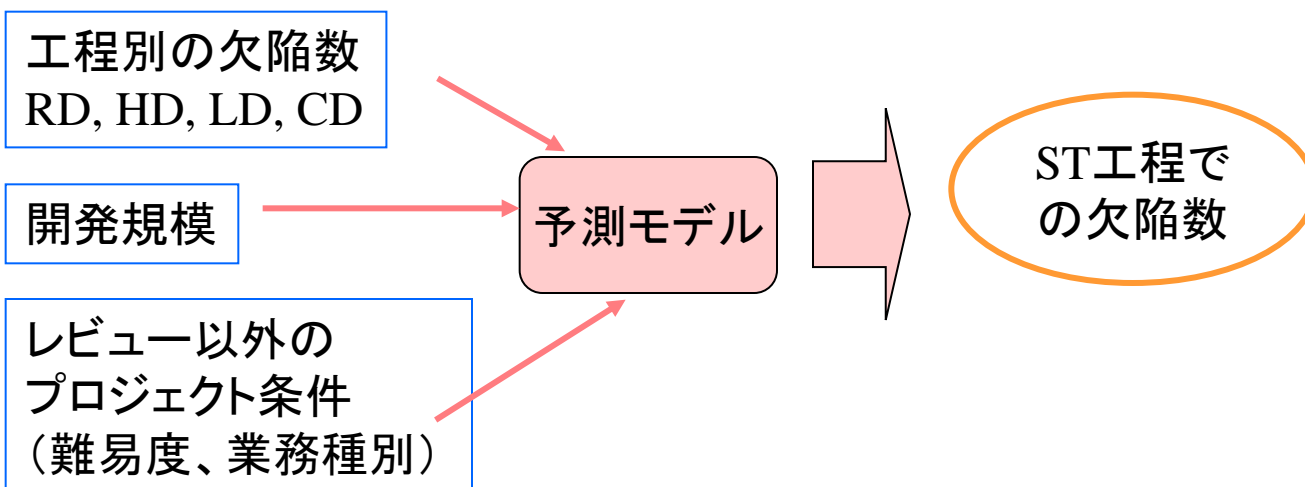


実施結果(イメージ)



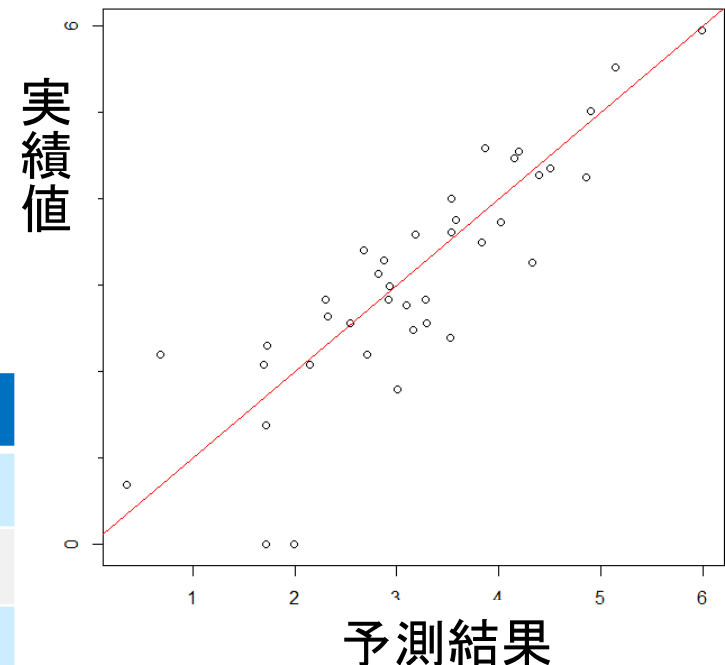
- プロジェクト実施途中での品質予測ができれば、先手管理により品質に起因するトラブルを予防できる。また、下流工程の欠陥数減少による工数削減効果も定量的に予測できる
- 技術的課題：総欠陥数の使用を避けて、(ほぼ)同じ指標値を実現したい
前倒し率 = (工程p以前に摘出した欠陥数) / (総欠陥数)
- 解決のアイデア：総欠陥数は開発規模の定数倍で近似できる
⇒ 総欠陥数を開発規模の定数倍で置き換えたモデルを構築

$$\begin{aligned} \log(\text{ST工程の欠陥数}) &= A + B \cdot \log(\text{開発規模}) \\ &+ C \cdot (\text{工程p以前の欠陥密度の和}) + D \cdot \log(\text{改造規模}) \\ &+ E_{\text{業務種別}} \cdot (\text{業務種別}) \end{aligned}$$



- 重回帰分析(ポアソン分布を仮定した一般化線形回帰)で検証
 - ただし、過分散が観察されたため、quasipoissonを指定(R言語, glmパッケージを使用)
- 対象工程 p= LD, CD, or UT
- 47件のプロジェクトデータ
- 目的変数 ST工程の欠陥摘出件数 (log をリンク関数として指定)
- 説明変数
 - 開発規模(の対数)
 - $d(p) = \sum_{q \leq p} f(q) / \text{size}$
 - 業務種別を表すダミー変数
 - 改造規模(の対数): 難易度をあらわす
 - 外注に掛けた工数の割合
- 擬似決定係数、p値など

予測結果と実績値(LD)



モデル対象工程	擬似R^2	F(p)の係数	F(p)のP値
LD	85.4%	-0.069	0.9%
CD	83.7%	-0.039	5.0%
UT	83.9%	-0.035	0.1%未満

● 効果説明モデル

- 「現実的で実態にあった品質計画策定が可能となり、品質向上できた」
- 「改善施策(候補)の効果を的確に見積もれるようになり、組織の改善施策の選定に役立っている」

● 品質予測モデル

- 当初、各工程の終了判定に使おうとしたが、あまり意味がなかった
 - ◆ 品質計画の値とほぼ同じになるケースが多かったため
- 各工程途中での中間予測結果を出して、(レビュー進展に伴う)品質向上の様子が見えるようにして、品質向上の動機づけになるようにした
 - ◆ その時点で当該工程を完了したとした時の予測値を算出

● 2つのモデルを適用してプロジェクト完了までいたった6プロジェクトの結果

- 5プロジェクトがベースライン値より品質向上(欠陥密度が半減:下表)
- 1プロジェクトは却って悪くなった
 - ◆ 悪化の原因は未経験の新規技術に挑戦したこと
 - ◆ 計画時に対応策を検討していたが結果的に十分でなかった

項目	Min	Max	中央値
開発規模(Kstep)	15	59	31
前倒し率 F(CD)の増加	3%	17%	15%
ST工程での欠陥密度減少	17%	93%	46%

- プロセス改善の原則：品質の高いプロセスが高い製品品質を導く
- プロセス品質を測る尺度
 - 工程別欠陥検出数割合
 - 前倒し率、累積前倒し率
- 工程ごとの欠陥件数間に定量的関係（べき乗減衰則）を仮定することで、2つの定量的モデルを構築した
 - 効果分析モデル
 - ◆ プロセス品質と製品品質の関係を明らかにしたモデル
 - ◆ 目標値策定等の応用がある
 - 品質予測モデル
 - ◆ 開発中の実績データ（欠陥件数等）からシステムテスト工程の欠陥数を予測するモデル
- どちらも実プロジェクトに適用し、品質管理に活用されている