

本スライドは、当日のセミナー資料の一部を抜粋したものです。

ソフトウェアメトリクスの測定 精度 の課題にどう対処するか

～測定精度を考慮した統計分析の実践～

ヤマハ(株)
小池 利和

アジェンダ

- ・ ソフトウェアメトリクス分析に統計手法を使う意義
- ・ 2群の母平均の差のt検定の解説
- ・ リリース後品質を予測するメトリクス
～t検定を実務で活用した事例～
- ・ レビューの改善効果を定量的に検証
～分散分析 & 多重比較法の解説～

ソフトウェアメトリクスでの統計手法活用の意義

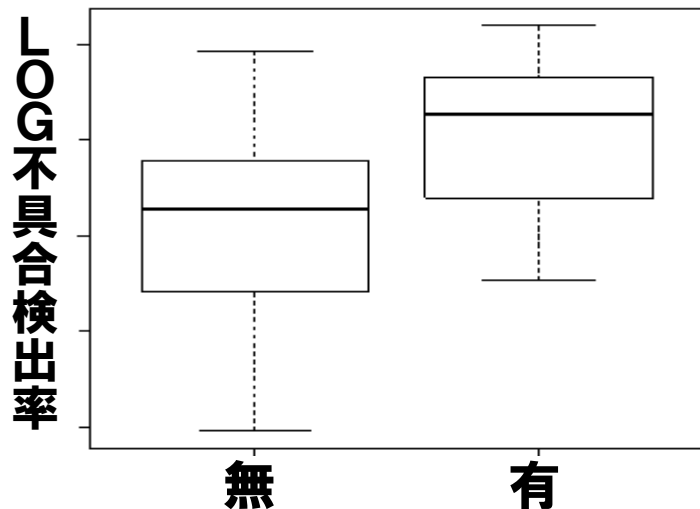
- 前頁でソフトウェアメトリクスは測定精度が悪いとありました。では、精度が悪いデータは本当に使えないのでしょうか？
- **ソフトウェアメトリクスは自然科学分野のデータ(長さ、重さなど)に比べて、測定誤差が大きいのは事実です。**
- しかしながら、**経済学や心理学など人間の行動を研究対象とする社会科学分野では測定データの誤差が大きいのは当たり前のこととして、調査結果から結論を導くのに統計手法の活用は必須**となっています。
- ソフトウェアメトリクスは人間の活動を測定している側面が大きいので、社会科学分野のデータに近い性質を持ちます。
- 精度が悪いから使えないと嘆くのではなく、だからこそ、**統計手法を駆使して、誤差に埋もれない重要な情報を読み取る必要がある**のです。

分析目的と使用メトリクス

- ・ プロジェクト完了時に算出していた以下の2つのメトリクスが製品リリース後の品質に関連があることを検証すると同時にどちらの方が関連が強く品質指標として妥当かを確認する。
 - 不具合密度 = ST不具合数 / 新規開発行数
 - 不具合検出率 = ST不具合数 / ST工数
- ※ST: システムテスト
- ・ 元々は、不具合密度を品質指標としていましたが、以下の理由で妥当ではないと感じていたことも背景となっています。
 - 不具合がテストが不十分で低いのか、品質が良くて低いかが区別できない
 - 信頼性工学では、MTBFなど時間と不具合の関係を見ることが一般的

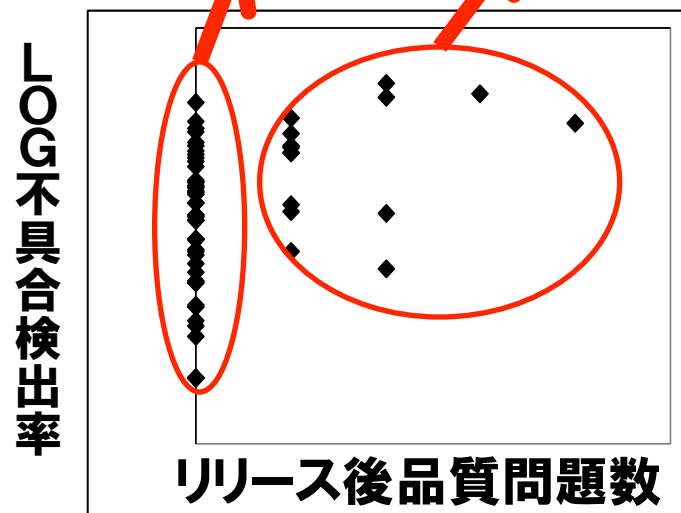
(補足) 相関分析と母平均の差のt検定の違い

品質問題有無の箱ひげ図



- **母平均の差のt検定**はデータ群で比較するので、多少の誤差によるデータ変動では結果は変わりません。つまり、鈍感(**ロバスト**)です。

誤差の多いソフトウェアメトリクスはよりロバストなアプローチを試みる



- **相関分析**は個々のデータの動きを見るので、誤差などのデータ変動に**敏感**です。

レビュー技法の効果を検証

例題)

以下にレビュー指摘による欠陥密度データが11個あります。データをそれぞれ適用したレビュー技法ごとに分類しています。レビュー対象の品質を考慮して技法を使い分けすることは特にしていない前提とします。

レビュー技法の違いは、欠陥密度に対して影響を与えているといえるでしょうか。

レビュー技法	レビュー指摘欠陥密度(件/頁)				平均
A(アドホックレビュー)	2.2	3.4	2.8	1.2	2.4
W(ウォークスルー)	3.0	4.4	5.8	—	4.4
I(インスペクション)	4.6	4.2	4.0	4.8	4.4
※アドホックレビューとは、技法を使わないレビューのこと	全体平均				3.7