

プロジェクトのデータ解析と 見積り研究会の紹介

(略称: PDA研究会)

2017年3月1日

プロジェクトのデータ解析と見積り研究会概要

プロジェクトの規模、工数、コスト、工期、品質、リスクなどの測定量を正しく分析し活用するためのデータ解析の手法を学び見積りおよびプロジェクト計画への活用を目的として研究する。

.背景・目的とそのフォーカス分野:

プロジェクトの規模、工数、コスト、工期、品質、リスクなどの、プロジェクトのマネジメントに関係する諸量を定量的に把握し、見積りおよびプロジェクト計画に活用したいという要望が増大している。
しかしながら、プロジェクトの測定量を正しく分析し活用するためには、データ解析の知識が必須となる。また、見積りに関しては様々な理論やツールがあり、これらのすべてを理解し活用するのは容易ではない。
本研究会は上記の状況に対応し、ITプロジェクトのデータ解析と見積り精度の向上、プロジェクト計画への活用を目的として学習、研究を行い、プロジェクト担当者が容易に活用できることに留意し、その成果を具体的に示すことでプロジェクトマネジメントに貢献することをめざす。

活動概要

- | | |
|--------------------|-----|
| 1. 統計活用セミナー(基礎的学習) | 完了 |
| 2. Excel統計(書籍活用研究) | 完了 |
| 3. CEBOKの活用研究 | 完了 |
| 4. 要求の仕様化技術の活用研究 | 継続中 |
| 5. R言語の学習と活用 | 継続中 |
| 6. アナリティクス手法の学習と活用 | 継続中 |
| 7. ソフトウェアメトリクス統計分析 | 継続中 |
| 8. QCDDRの予測と管理 | 継続中 |
| 9. 論文投稿 | 継続中 |

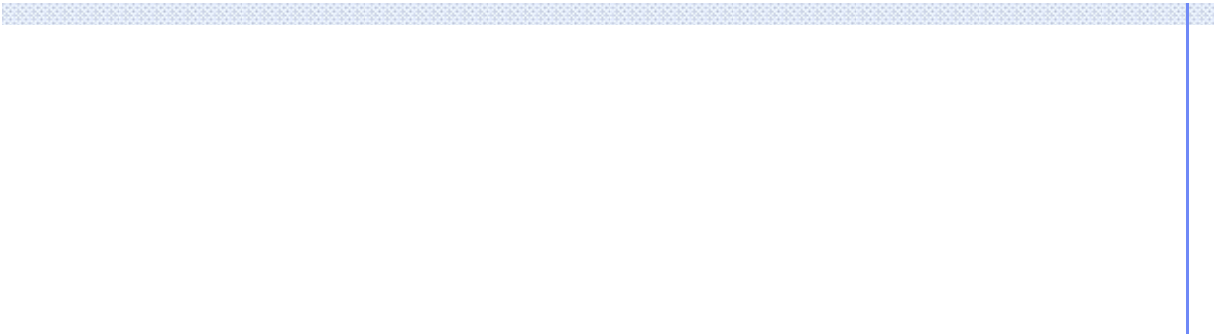
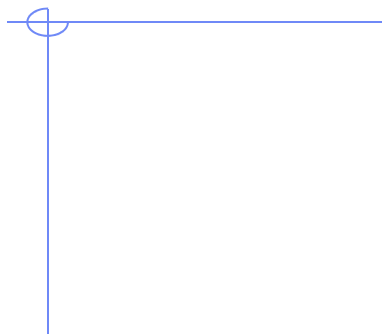
活動形態

1. 月1回の会合（参加企業の会議室などを利用）
2. 統計の基礎から応用までを学習
3. 実践に活用可能なコンテンツの開発
4. 学習の成果または研究成果発表の場を提供

参加者はメンバーが開発した各種資料を活用できます。

Excel統計(書籍活用研究)

No.	学習内容	書籍1: らくらく統計解析		書籍2: Excel関数		書籍3: 回帰分析		書籍4: 図解統計学	
		解説	関数またはアドイン	解説	関数またはアドイン	解説	関数またはアドイン	解説	関数またはアドイン
1	統計的方法活用への道							○	
2	確率の初歩	*		*				○	
3	確率と期待値	○						○	
4	事象と確率変数	○						○	
5	条件付き確率(ベイズの定理)	○	累積度数グラフ					○	
6	平均とバラツキ	○	AVERAGE	○	AVERAGE	○	分析ツール	○	
7	平均値	○		○		○	分析ツール	○	
8	加重平均と幾何平均						分析ツール		
9	分散と標準偏差	○	正規分布	○	VAR,STDEV	○	分析ツール	○	
10	変動係数								
11	平均値と分散の計算	○	正規分布	○		○	分析ツール	○	
12	属性の数量化					○	分析ツール	○	
13	最大値と最小値	○	MAX、MIN	○	MAX、MIN	○	分析ツール	○	
14	中央値とレンジ	○	MEDIAN	○	MEDIAN	○	分析ツール	○	
15	最頻値(モード)	○	MODE2(アドイン)	○	MODE	○	分析ツール	○	
16	歪度と尖度(ひずみ、とがり)			○	SKEW,KURT				
17	度数分布	*		*				○	
18	ヒストグラム	○	ヒストグラム		FREQUENCY	○	分析ツール	○	
19	幹葉図				ドットプロット				
20	累積度数グラフ	○	累積度数グラフ						
21	パーセント点			○				○	
22	相関	*		*				○	
23	相関・回帰の意味(直値)			○				○	
24	相関・回帰の意味(対数変換)								
25	擬似相関							○	
26	相関図と相関係数	○	散布図	○	CORREL	○	分析ツール	○	
27	相関係数の計算	○		○				○	
28	共分散	○						○	
29	相関係数についての注意								
	クロス集計表							○	
31	母集団と標本	*		*				○	
32	母集団と標本	○		○				○	
33	離散分布からの標本抽出	*		*				○	
34	ベルヌーイ分布							○	
35	2項分布							○	
36	2項分布からの標本抽出							○	
37	割合の期待値と分散							○	
38	ポアソン分布							○	
39	ポアソン分布の期待値と分散							○	



統計活用セミナー

ここでは、一部を紹介します。

統計的方法～活用への道

●教材内容の概略

I. 手法の案内

どの分野の方にも理解しやすいように健康に関する例題を載せます。(多変量解析へのガイド)

付8 不揃いなデータにおける分散分析の解析例

付9 統計的解析に役立つデータ表をつくる

II. 調査計画の演習

III. 分布に馴染むための手作業

付1 $N(0, 1)$ の正規乱数

付2 正規確率紙の使い方とパーセントイル法

付3 BASICプログラム

正規乱数・乱数の発生・＜独立試行／幾何分布／ポアソン分布／指数分布＞のシミュレーション・二項分布の確率

付4 説明補足

幾何分布の期待値

ポアソン分布と指数分布

母出現率の推定

調査で副作用を検出する確率

副作用調査の必要症例数

付5 乱数表

付6 正規確率紙

付7 解析法により判定が異なることがある計算例

【本講座では、学習の進捗を自分でチェックできます】

- ・なぜ統計的方法が必要か？・差の大きさは？
- ・なぜ「ばらつき」が生じるか？
- ・標準偏差(SD)と標準誤差(SE)の意味は？
- ・外れ値の扱いは？
- ・電卓キーにある $\sigma n, \sigma n-1$ の使い分は？
- ・ランダム化の意味は？
- ・実験例数(N)の決め方は？
- ・なぜ正規分布に従うか？
- ・検定の切れ目を良くするには？
- ・正規分布の確認は？
- ・切れ目と普遍性の両立の条件は？
- ・対数正規分布の場合は？
- ・不偏推定値とは？
- ・変数変換による影響は？
- ・信頼度とは？危険率とは？
- ・正常域の決め方は？
- ・両側、片側仮説の使い分けは？
- ・平均値を何に使うか？
- ・相関と因果関係の確認の方法は？
- ・幾何平均を何に使うか？

どれくらいわかりますか？

芝浦工業大学「オープンテクノカレッジ」より引用

家族に関する俗説(1/3)

男女はほとんど同じ確率で生まれるのだから、女の子である確率は $1/2$ に違いない。

ブラウンさんには二人の子がいます。そのうち少なくとも一人は男の子です。もう一人の子が女の子である確率はいくつでしょう。



スティーブン・セン「確率と統計のパラドックス」より引用

家族に関する俗説(2/3)

確率は同じなのに、
もらうほうが
多いので賭ける
方が得！

男の子だったら 5万円払います。
女の子だったら 4万円いただきます。
賭けますか？



家族に関する俗説 (3/3)

ほとんど人が、女の子である
確率は $1/2$ と考えます。
しかし、正しい確率は $2/3$ です。

信じられない！



CEBoKの活用研究

日本コスト評価学会 (JSCEA) の
CEBoK (Cost Estimating Body of Knowledge) を活用

信頼区間

JSCEA CEBOK より引用

- ◆ b に対する95%信頼区間を計算するための方法を、問題の例を使用して説明します:



例:

$$\hat{Y} = 2.5 + 0.6 X$$

$\alpha = 0.05$
 $df = 5 - 2 = 3$

信頼区間:

$$= \hat{b} \pm t_{\alpha/2, df} \times \text{stdev}$$

$$= 0.6 \pm t_{.025, 3} \times 0.3$$

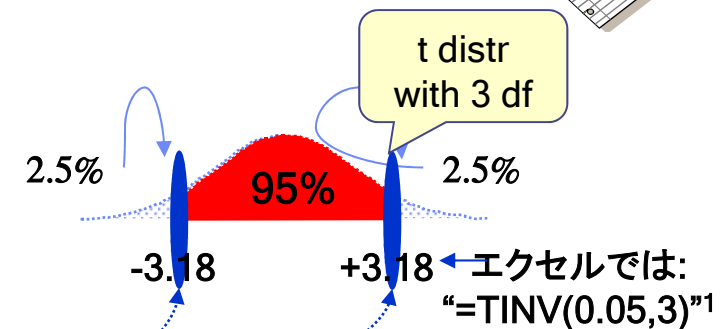
$$= 0.6 \pm 3.18 \times 0.3$$

$$= (-0.3, 1.4)$$

このようにして、95%の信頼度でbの真の値がこの区間に含まれていると言う事が出来る。

注1: エクセルのTINV()関数は、入力された値が両端に分かれていると仮定している為、 $\alpha/2$ の代わりに α を入力する必要がある。

ヒント: この結果は、直接、t(テスト)とx(係数)に関連している。もしも、信頼区間がゼロを含む場合、係数は統計的に有意ではないと言える。



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.79
R Square	0.62
Adjusted R Square	0.50
Standard Error	2.5
Observations	5

エクセルではそれぞれのパラメーターの95%信頼区間を計算してくれる。

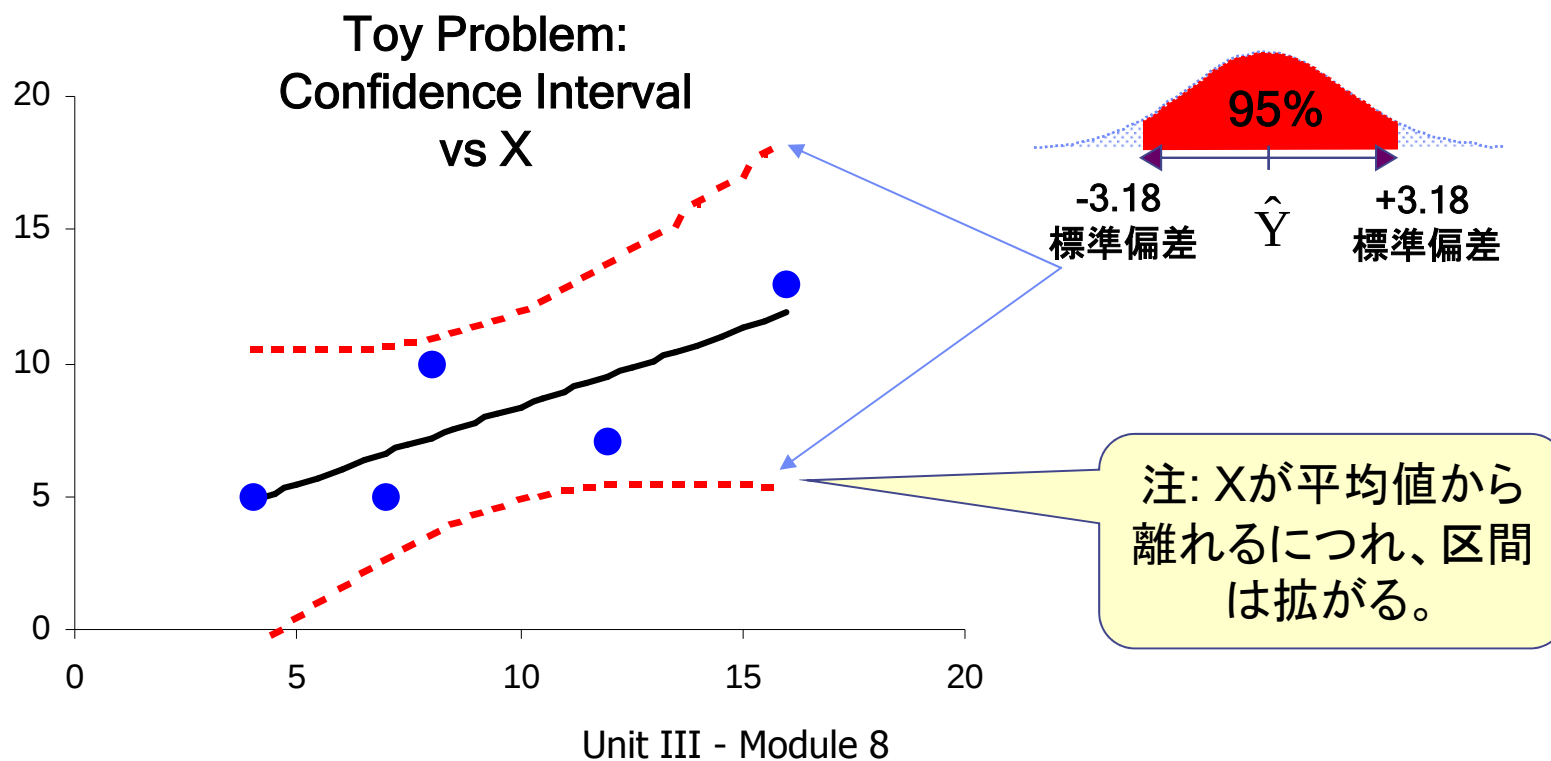
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	29.8	29.8	4.9	0.11
Residual	3	18.2	6.1		
Total	4	48			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2.5	2.7	0.9	0.42	-6.1	11.1
x	0.6	0.3	2.2	0.11	-0.3	1.4

信頼帯

JSCEA CEBok より引用

- ◆ 全てのXに対する信頼区間をプロットする時、“信頼帯”と呼ぶ、2本の双曲線が描かれる。



予測区間

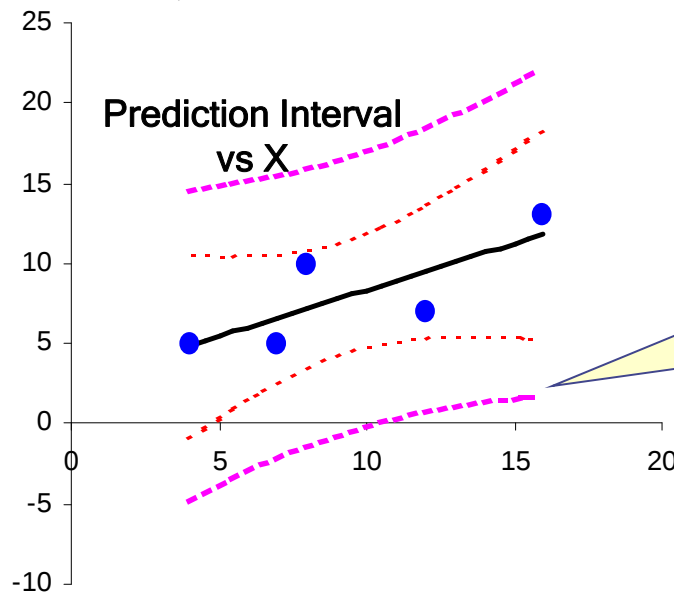
JSCEA CEBok より引用

- ◆ Xの値が与えられた時の、Yの信頼区間を新たに知りたい時がしばしばあります。
 - これは予測区間と呼ばれ、数式は以下の通りです:

$$\hat{Y} \pm t_{\alpha/2, df} \times \text{SEE} \sqrt{\frac{n+1}{n} + \frac{(X - \bar{X})^2}{\sum X^2 - n\bar{X}^2}}$$

Xを固定した時の
Yの標準偏差

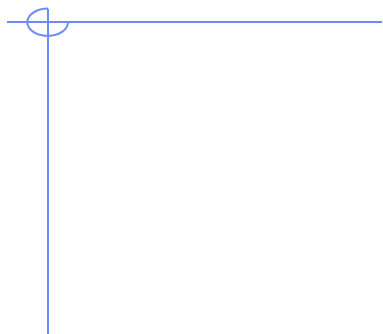
tの分散は、以前計算
したのと同じ棄却限
界値を持つ



ヒント: 予測区間はy
ハットに対比して、y
の区間を示す

予測区間は、信頼区間よ
りも幅が広い事に注意

Unit III - Module 8



要求の仕様化技術の活用研究

要求の仕様化技術であるUSDM (Universal Specification Describing Manner) で記述された仕様書をソフトウェアの規模計測に活用する方法を研究

話題沸騰ポットの計測

一つの製品では仕様の複雑さに伴うメトリクスの増加を計測できない！

そこで…

話題沸騰ポットの仕様の仕様から、以下の3種類の製品の仕様を定義する。

- ① 低機能ポット … 基本的な機能のみを持つポット
- ② 中機能ポット … ロック解除、蓋センサーの機能を追加
- ③ 高機能ポット … キッチンタイマー、3段階保温温度の機能を追加

USDMに従った仕様書

USDM		
pot-210		コンセントの抜き差しで、ポットを利用できない状態/利用できる状態にする。
	pot-210-11	コンセントを差し込むと、設定値にはデフォルト値がセットされ、ポットが機能する状態(アイドル)になる。
	pot-210-12	コンセントを抜くと、ポットは蓋の開け閉め以外は何も機能しなくなる。
pot-220		アイドルの状態、蓋を閉じたら、水位を確認し、条件に合えば沸騰行為をする。
	pot-220-11	蓋センサーが3sec以上onとなったら、蓋が閉じられたと判断する。
	pot-220-21	蓋が閉じられ、水量が適正な場合、沸騰行為をする(A)。
	pot-220-31	蓋が閉じられても、水量が異常な場合、状態はアイドルのままである。

USDMに従った仕様書は記述の粒度が均一になる

論文投稿

1. PM学会2011年度秋季研究発表大会

プロジェクトの見える化と予測に基づく定量的管理の実現

2. PM学会誌(2011年10月号)解説記事

ソフトウェア開発プロジェクトの計数管理フレームワークによる定量的管理

3. PM学会2014年度春季研究発表大会

CEBoKを活用したコストとスケジュールのリスク分析

4. PM学会2016年度春季研究発表大会

要求の仕様化技術に伴うメトリクスの実用と活用

参加者は論文や記事の詳細を知ることができます。

メンバー

No	氏名	所属	備考
1	梶山 昌之	株式会社ワイハット	主査
2	楠本 真二	大阪大学	副主査
3	小川 純平	株式会社 日立製作所	幹事
4	藤井 英章	日本アイ・ビー・エム株式会社	
5	吉田 賢吾	三菱電機インフォメーションシステムズ株式会社	
6	大岩 佐和子	財団法人経済調査会	
7	押野 智樹	財団法人経済調査会	
8	安藤 也寸子	日本アイ・ビー・エム株式会社	
9	鈴木 圭一	富士フィルム	
10	柏 知	日本アイ・ビー・エム株式会社	
11	藤木 淳夫	株式会社DTS	
12	石井 祐介	株式会社DTS	
13	岡野 麻子	三菱スペース・ソフトウェア株式会社	
14	馬崎 哲	シュルンベルジェ株式会社	

募集中！

データ解析の実力が
身につく研究会です。



参加者募集中

参加希望の方は下記までご連絡ください。
梶山昌之 kajiyama@yhat.co.jp

募集中！

データ解析の実力が
身につく研究会です。

