

2024年2月22日  
データサイエンティスト集会 in VRC

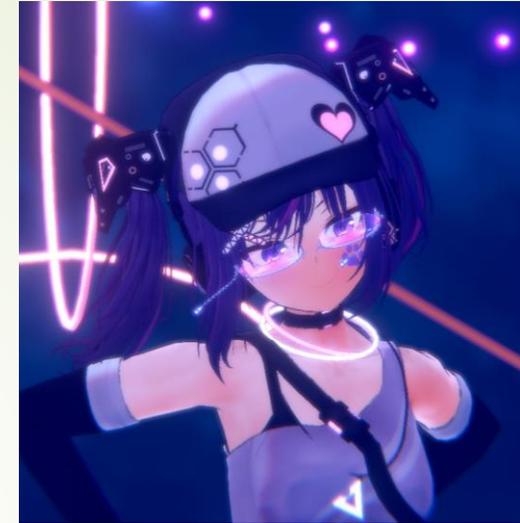
DS初心者向け  
知名度は低いけどオススメの良書紹介 1  
～少量データの取り扱い～

1

ぶんちん

# 自己紹介 ぶんちゃん

- ▶ 複合経営が特徴の企業（製造業）に所属
- ▶ データ分析担当者だったが。。。



良い書籍はたくさんあるけど。。。。

▶ 良書は人気になりやすい



知られていない本でも良書がある  
今回はそれを紹介

# 本当にデータ量が少ない領域

世間の話題

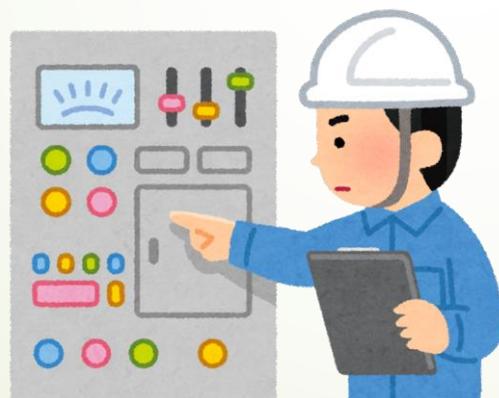
- 大量のデータを有効活用！
- これからは DX！ AI！



少量データ  
N=1000以上



実験系



新製品・新設備

データ量は1～2桁  
そんなにデータ量ない  
どうしたら良いの？



でも、成果は求められる。当然だよね？

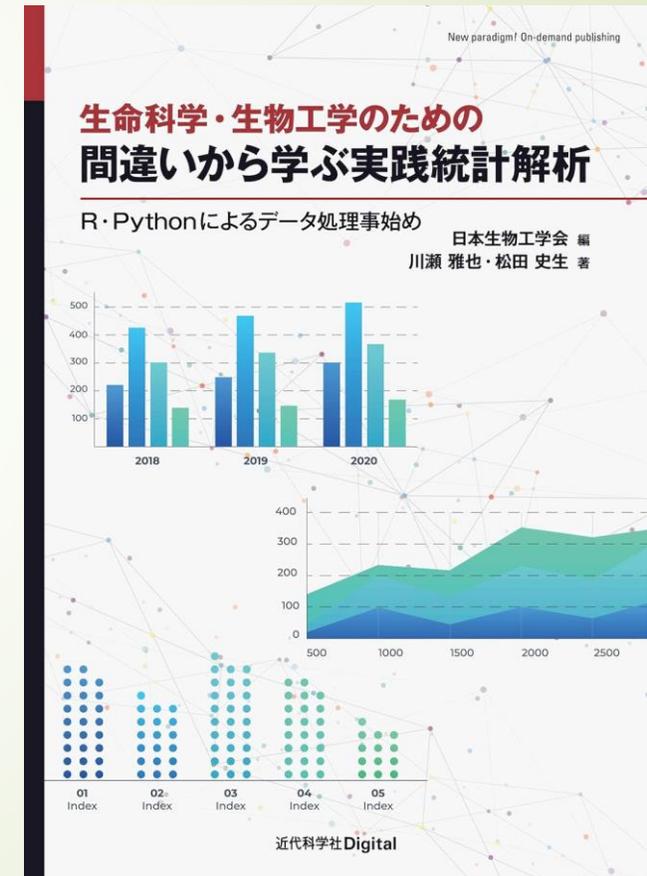
# 生命科学・生物工学のための 間違いから学ぶ実践統計解析

ここで興味を持った人も、  
購入はちょっと待って！

- ▶ 現実的な範囲で適切に統計学を使いこなそう
- ▶ データ収集コストが重い実験系のデータの取り扱いについて説明した書籍
- ▶ 元は日本生物工学会の学会誌に掲載されていた連載記事



学生2名に先生が指導の形式



# 目次

- ▶ 第1部 Rを使って統計解析を行おう
  - ▶ 第1章 平均値にご注意を
  - ▶ 第2章 正規分布を極める
  - ▶ **第3章 データの数はいくつ必要？**
  - ▶ 第4章 平均の差の検定の使い方
  - ▶ 第5章 正しい統計記述とは？
  - ▶ **第6章 外れ値にご用心**
  - ▶ **第7章 多重比較って何？**
  - ▶ 第8章  $\chi^2$ 検定の使い方？
  - ▶ 第9章 相関と相関係数
  - ▶ **第10章 単回帰は難しい**
  - ▶ 第11章 誤差の伝播
  - ▶ 第12章 直交表と重回帰分析
- ▶ 第2部 Pythonでも統計解析を行えるようになろう
  - ▶ 第13章 Python？
  - ▶ 第14章 Pythonの文法 分岐と繰り返し
  - ▶ 第15章 Pythonによる統計入門1
  - ▶ 第16章 Pythonによる統計入門2
  - ▶ 第17章 主成分分析その1, 方法のおさらい
  - ▶ 第18章 主成分分析その2, 結果を解釈する
  - ▶ **第19章 偽反復**
  - ▶ 第20章 階層クラスター分析はちょっときまぐれ
  - ▶ **第21章 微妙な時のしきい値が肝心**
  - ▶ 第22章 深層学習, すぐできます
- ▶ 第3部 統計解析の基本を見直そう
  - ▶ **第23章 p値とサンプルサイズ**
  - ▶ **第24章 統計処理の落とし穴**

# 第10章 単回帰は難しい

- 思考停止で単回帰モデルをつくっていませんか？
- まさかR2だけを見ているだけなんてないですよ？
- 明確な答えはないため、考えましょう。

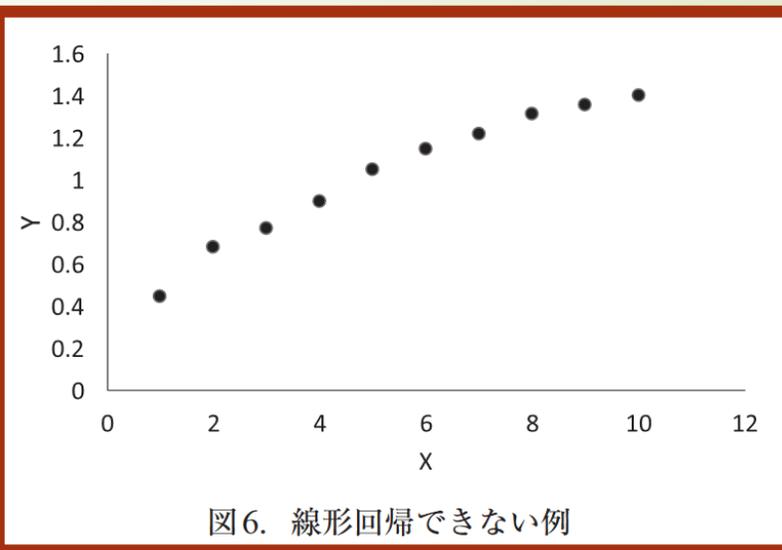


図6. 線形回帰できない例

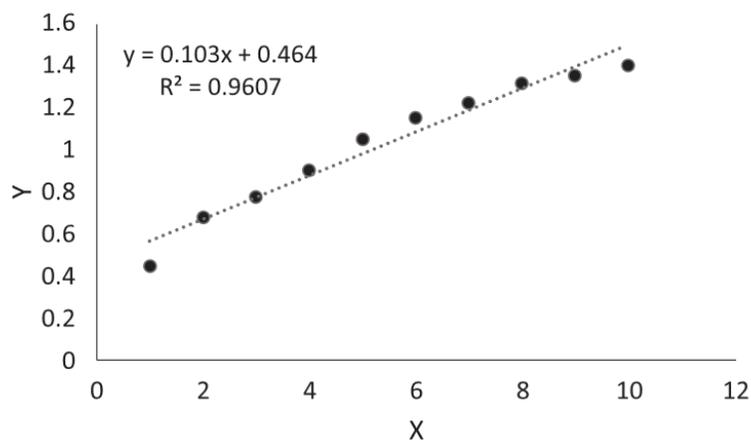


図7. 線形回帰できない例（線形回帰）

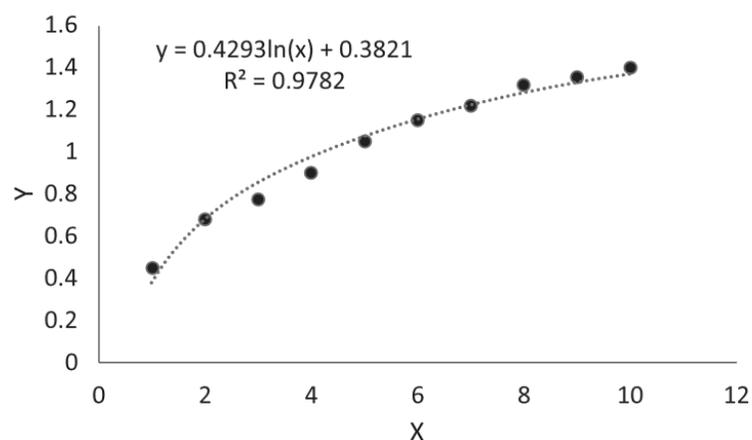


図8. 図7を対数回帰したケース

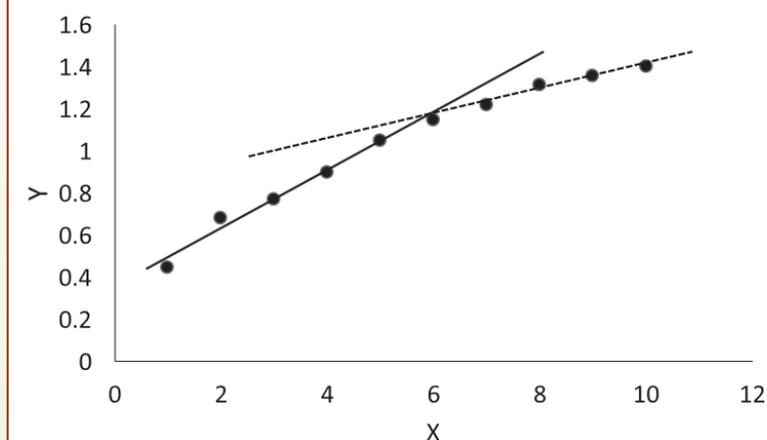


図9. 図7を2直線で線形回帰したケース

# 初心者が直面する壁

将来のためにデータサイエンスの勉強するぞ！  
統計！機械学習！プログラミング！etc



勉強したことが仕事で使えない。  
使ってみたけど成果につながらない。  
実力不足。もっと勉強しないと！



Q.いつになったら  
仕事に生かせるの??  
成果を出せるの??

**A.個別事例ごとに"考える"ことが重要  
その練習をするのが本書**

# 一般公開の資料

- ▶ 「生命科学・生物工学のための間違いから学ぶ実践統計解析」で検索すると右のページが出てくる
- ▶ 実は日本生物工学会のHPで全記事が公開済
- ▶ 自分用だけではなく、誰かの指導の参考にもおすすめ
- ▶ 近々、社内向けにこの本を使った講義を予定

The Society for Biotechnology, Japan  
公益社団法人  
日本生物工学会

ホーム > 生物工学会誌 > 生物工学会誌 - 『間違いから学ぶ実践統計解析』 (2016年94巻4号~2018年96巻2号 隔月掲載)

生物工学会誌 - 『間違いから学ぶ実践統計解析』 (2016年94巻4号~2018年96巻2号 隔月掲載)

投稿日: 2017.12.26 最終更新日: 2020.02.20

いいね! 0 ツイート

こちらでは、生物工学会誌のシリーズ企画『間違いから学ぶ実践統計解析』 (2016年94巻4号~2018年96巻2号 隔月掲載) のPDF版をご覧いただけます。⇒『[続・間違いから学ぶ実践統計解析](#)』はこちら

掲載記事	著者	巻号-頁 (掲載年)
第1回 平均値にご注意を	川瀬 雅也 松田 史生	94-4-208 (2016)
第2回 正規分布を極める	川瀬 雅也 松田 史生	94-6-348 (2016)
第3回 データ数はいくつ必要	川瀬 雅也 松田 史生	94-8-510 (2016)
第4回 平均の差の検定の使い方	川瀬 雅也 松田 史生	94-10-656 (2016)
第5回 正しい統計記述とは?	川瀬 雅也 松田 史生	94-12-788 (2016)
第6回 外れ値にご用心	川瀬 雅也 松田 史生	95-2- 96 (2017)
第7回 多重比較って何?	川瀬 雅也 松田 史生	95-4-214 (2017)
第8回 $\chi^2$ 検定の使い方?	川瀬 雅也 松田 史生	95-6-348 (2017)
第9回 相関と相関係数	川瀬 雅也 松田 史生	95-8-494 (2017)
第10回 単回帰分析はむずかしい	川瀬 雅也 松田 史生	95-10-624 (2017)
第11回 誤差の伝播	川瀬 雅也 松田 史生	95-12-752 (2017)
第12回 直交表と重回帰分析 ⇒mlr.csv	川瀬 雅也 松田 史生	96-2-88 (2018)

⇒過去号掲載記事 (記事種別) 一覧へ  
⇒生物工学会誌Topへ

投稿者 学会事務局