

統計的有意差(p値)を巡る最近の論争： Moving to a World Beyond “ $p < 0.05$ ”

慶應義塾大学大学院

中村知繁

tomoshige.nakamura@gmail.com

慶應義塾大学三田キャンパス

Updated on May 14th, 2019

目次.

- ▶ インTRODクシヨン
 - ▶ この発表で扱う内容と、発表のサマリー
- ▶ p値 < 0.05の罫。
 - ▶ p値とは何か？
 - ▶ p値を使う罫は「足元の想定」である（ここがポイント！）
 - ▶ p値に対するASAのPrinciple
- ▶ p値の問題を受けて、何をすべきか（Moving to a world beyond “p < 0.05”）
 - ▶ 「統計的有意」からの脱却
 - ▶ ATOMという原則.
- ▶ 現実的なガイドラインとして
 - ▶ 私たちが、「p < 0.05の世界」から脱却するために

目次.

- ▶ インTRODクシヨン
 - ▶ この発表で扱う内容と、発表のサマリー
- ▶ p 値 < 0.05 の罟。
 - ▶ p 値とは何か？
 - ▶ p 値を使う罟は「足元の想定」である（ここがポイント！）
 - ▶ p 値に対するASAのPrinciple
- ▶ p 値の問題を受けて、何をすべきか（Moving to a world beyond “ $p < 0.05$ ”）
 - ▶ 「統計的有意」からの脱却
 - ▶ ATOMという原則.
- ▶ 現実的なガイドラインとして
 - ▶ 私たちが、「 $p < 0.05$ の世界」から脱却するために

このプレゼンテーションで扱う内容



- ▶ このプレゼンテーションでは、以下の論文誌の2019年の特集にある 次の論文を扱います
 - ▶ 論文誌：The American Statistician
 - ▶ Volume : 73
 - ▶ 特集：Statistical Inference in the 21st Century: A World Beyond $p < 0.05$
 - ▶ 論文：Wasserstein. R, Schirm. A, and Lazar. N. (2019). **Moving to a World Beyond “ $p < 0.05$ ”**
- ▶ 論文の内容はインターネットでアクセスすれば閲覧可能なので、詳細な内容については、ご自身でお確かめください。
- ▶ この発表では、近年行われている「p値についての議論」を30分でコンパクトに説明することを目標とします。

このプレゼンテーションで扱う内容



COMMENT • 20 MARCH 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein , Sander Greenland & Blake McShane

- ▶ また、本発表の内容は2019年3月にNatureで発表された上記の記事の内容を含みます。
- ▶ この記事も、学術的な研究においてP値の不適切な使用（主に、P値を仮説の立証の根拠とすること）に対しての警鐘を鳴らすものです。

* 注

- ▶ スライドでは、p値についての正しい理解のために、改めてp値とはなにかについて記載したスライドがありますが、当日具体的に説明する予定はありません。
- ▶ 今回紹介する論文でのp値誤用問題の想定は学術的な研究ですが、今回のイベントへお越しの皆様の中には、ビジネスで使用する方々もいらっしゃると思います。よって、論文の内容をそのまま話すのではなく、必要な箇所のみ抽出して、発表者の意見を加えて記述しています。
- ▶ 以上の点に、ご了承頂ますようお願いいたします。

発表のサマリー - I

ここ数年の流れ

- ▶ 2014年に心理学系の論文誌Basic and Applied Social Psychologyで、p値の使用が禁止された。
- ▶ そのあと、Nature, ScienceNews, Statisticianなどで、p値に基づく科学的な論文は、再現性の保証などの点から警鐘が鳴らされました。
- ▶ 2016年には、ASA（アメリカ統計学会）がp値に関する声明を発表。p値の使用に関する6つの原則を提示。
- ▶ 2019年3月のAmerican Statistianでは「 $p < 0.05$ から脱却し、データ解析は次のステージへ」という特集が組まれる。

論文の概略

- ▶ p値を用いた二元論的な発想から脱却しよう！
 - ▶ 「統計的に有意」という言葉を捨てよう！
- ▶ p値を理由にした「結論」をやめよう！
 - ▶ p値が小さくても関連がない場合もあれば、p値が大きくても関連がある場合もある。
 - ▶ p値は、学術的に意味があることと常に同義ではない

発表のサマリー - II

論文の概略

- ▶ 「統計的に有意」という言葉が果たしてきた役割に変わるものは存在しないが、私たちは次の4つを意識して行動していく必要がある。
 - ▶ Accept Uncertainty (データ解析の不確実性を受け入れること)
 - ▶ Be Thoughtful (データの解析に対して慎重で、思慮深くあること)
 - ▶ Be Open (データの解析のプロセスや想定などを公開すること)
 - ▶ Be Modest (データの解析の結果について謙虚であること)
- ▶ データの解析を評価する際のポイントも変えていく必要がある。
 - ▶ 得られたデータ解析の結果は、再現可能なものか (研究目的に対して、デザインが適切になされているか、また手法の選択のロジックは適切か)
 - ▶ データ解析の価値は「結果」にあるのではない。解析を行った「前提」や、その過程で置かれた「仮説」は十分に議論されたものか。
 - ▶ 例えば、モデル選択で盲目的にStepwise Regressionをしていたら、それはモデルに対しての配慮が明らかに足りないと思って良い。

発表のサマリー - III

私たちはどうするか

- ▶ この特集号に寄稿しているのは、ほとんどが研究者なので、内容のスポットは特に研究へと向けられている。
- ▶ 実際、治験などにおいてはp値の有用性については、柳川(2017)などで述べられている通り、状況次第である。
- ▶ しかしながら、「p値の誤用」と「統計的有意の乱用」は、データ解析を行った際のレポートで度々目にする。
- ▶ 例えば、回帰モデルにおける「 $p < 0.05$ 」や「AIC」の乱用がはその一例である。
- ▶ Stepwise法やL1-罰則正則化で正しいモデルが選択できているというような「過信」が蔓延している。
- ▶ 今こそ、データの取得から、モデルの構築、レポートの作成までの、データ解析のフローを見直し、自分たちのデータ解析に適した解析フローを作り直す必要がある。

Moving to a World Beyond “ $p < 0.05$ ”

目次.

- ▶ イントロダクション
 - ▶ この発表で扱う内容と、発表のサマリー
- ▶ p 値 < 0.05の罠。
 - ▶ p 値とは何か？
 - ▶ p 値を使う罠は「足元の想定」である（ここがポイント！）
 - ▶ p 値に対するASAのPrinciple
- ▶ p 値の問題を受けて、何をすべきか（Moving to a world beyond “ $p < 0.05$ ”）
 - ▶ 「統計的有意」からの脱却
 - ▶ ATOMという原則.
- ▶ 現実的なガイドラインとして（発表者の意見）
 - ▶ 私たちが、「 $p < 0.05$ の世界」から脱却するために

p値とは何か - I (仮説検定概説)

▶ 統計的仮説検定とは？

- ▶ 帰無仮説を棄却し対立仮説を支持するか，又は帰無仮説を棄却しないかを観測値に基づいて決めるための統計的手続き。
- ▶ その手続きは，帰無仮説が成立しているにもかかわらず棄却する確率が α 以下になるように決められる。この α を有意水準という。

▶ P値とは？

- ▶ **与えられた**データの値に対して、帰無仮説を棄却できる最小の有意水準

▶ P値とは？（噛み砕くと）

- ▶ P値とは、特定の統計モデルのもとで、データの統計的な要約（例えば2つの群の平均の差、回帰係数の推定値）が矛盾する程度を示す指標！
- ▶ 配布資料には、理論的なP値の定義が書かれている。

p値とは何か - II（仮説検定の例：回帰係数）

- ▶ Example：線形回帰モデル
- ▶ 例えば、データに線形回帰モデルを当てはめた場合の、回帰係数 β に対する検定では、 $H_0 : \beta = 0$ および $H_1 : \beta \neq 0$ である。
- ▶ このとき、有意水準5%の検定（ $p < 0.05$ で有意とする検定）は、帰無仮説が正しい（ $\beta=0$ ）なのに、 β は0ではないとしてしまう確率を5%以下に抑えるような検定を行っていることに対応する。
- ▶ 注意すべきこと
 - ▶ Type II Errorについては何も言及していない。
 - ▶ データ取得や、モデルの正しさについては一切言及されていない。
 - ▶ 影響の大きさについては一切言及しない。

p値を使う罣は「足元の想定」である - I

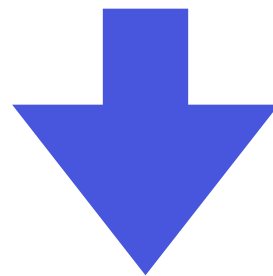
- ▶ 仮説検定の理論の結果は、一般的に
 - ▶ $X_1, \dots, X_N$ が統計モデル $\{f(x; \theta) ; \theta \in \Theta \subset \mathbf{R}\}$ からランダムな標本が得られたと仮定する。
- ▶ という魔法のことばが最初についている。

現実はどうだ！？

- ▶ データ解析では、次の2つがわからない。
 - ▶ データが、本当に知りたい母集団からランダムに取られているかどうか
 - ▶ データが、どんな分布やモデルから生成されたか
- ▶ 実際の解析では、魔法のことばの大前提から疑う必要がある。
- ▶ 推定量や、その検定は、これらについて十分な吟味と考察の上にしか意味を持たない。

p値を使う罣は「足元の想定」である - II

- ▶ 【良くない解析レポートの例】
 - ▶ データの取得方法・解析モデルの妥当性についての報告ができていない
 - ▶ P値や、特定の統計的な基準値が高いことを、判断の理由としている。
- ▶ このような論文は一定数存在していて、また業務におけるデータ解析のレポートでも散見される。



【検証】

P値はそんなに信用できるのか！？

p値を使う罣は「足元の想定」である - III (Exampleその1)

- ▶ 以下のモデルから、データを発生させる。

$$X_1 \sim N(0,1)$$

$$X_2 = X_1 + N(0,1)$$

$$X_3 \sim \text{Bernoulli}(\text{expit}(X_1 + X_2))$$

$$X_4 = |X_1 + X_3X_2|$$

- ▶ 結果変数に対するモデルは、

$$Y = 0.2X_2 + 0.2X_3 + N(0,1)$$

- ▶ このとき、サンプル数を200として、1000回繰り返してデータを発生させた。
- ▶ 発生させたデータに、すべてのXを説明変数とした線形回帰モデルを当てはめて、回帰係数を推定し、p値を計算して、0.05を下回った回数を、以下の表にまとめた。

	切片	X_1	X_2	X_3	X_4
有意になった回数	51	53	697	201	34

繰り返し回数1000回中、 $p=0.05$ で有意になった回数

p値を使う罣は「足元の想定」である - III (Exampleその2)

- ▶ 解析をしていて、 X_4 が効果があるような結果がほしいと思った！
- ▶ 変数を抜いてみよう。。。

N=200	切片	X_1	X_3	X_4
有意になった回数	150	322	433	115

- ▶ 結果変数に対するモデルに含まれない変数が有意になった・・・
- ▶ サンプル数を更に増やして、 $N = 1000$ にしたら

N=1000	切片	X_1	X_3	X_4
有意になった回数	630	921	978	381

- ▶ 簡単に $p < 0.05$ を下回る回数を増やすことができた。つまり、p値なんてモデルの仮定とサンプル数でいかようにもコントロール可能。 $N=5000$ はひどい。

N=5000	切片	X_1	X_3	X_4
有意になった回数	1000	1000	1000	973

有意にしたい変数は、有意にできる！
ビッグデータ最高ですね♡（皮肉）

p値に対するASAのprinciple (Wasserstein RL, Lazar NA. (2017))

1. P値は、データと自分が仮定した統計モデルの矛盾の程度を示す値である。
2. P値は、調べている仮説が正しい確率や、データが偶然のみでえられた確率を測るものではない。
3. 科学的な結論や、ビジネス/政策における決定は、P値がある閾値を超えたかどうかを根拠にするべきではない。
4. 適切な統計的推論のためには、研究の中で調べる仮説の数、データ収集の際に行ったすべての決定、実行したすべての統計解析、そして計算したすべてのP値を研究者は開示すべきである。
5. P値や統計的有意性は、効果の大きさや結果の重要性を意味しない。
6. P値は、それだけでは統計モデルや仮説に関するエビデンスの、よい指標とはならない。

目次.

- ▶ イントロダクション
 - ▶ この発表で扱う内容と、発表のサマリー
- ▶ p 値 < 0.05 の罠。
 - ▶ p 値とは何か？
 - ▶ p 値を使う罠は「足元の想定」である（ここがポイント！）
 - ▶ p 値に対するASAのPrinciple
- ▶ p 値の問題を受けて、何をすべきか（Moving to a world beyond “ $p < 0.05$ ”）
 - ▶ 「統計的有意」からの脱却
 - ▶ ATOMという原則.
- ▶ 現実的なガイドラインとして（発表者の意見）
 - ▶ 私たちが、「 $p < 0.05$ の世界」から脱却するために

統計的有意からの脱却 - I

P値誤用の流れ

- ▶ Edgeworth(1885)/Fisher(1923)ではP値は、結果の考察のための道具（P値の数値そのものを気にしていた）
- ▶ Neyman-Pearsonの統計的仮説検定が広まり、 $P < 0.05$ が生まれ「有意であること」や「Statistically Significant」という言葉生まれる。
- ▶ ある水準を下回った（有意）＝「意味がある」という謎の関係が生まれる。
- ▶ いまは、有意じゃなければ、publishしないという悪い風潮になり、科学的な論文で出版されるものは有意なものばかり・・・（出版バイアス）



統計的に有意



重要な結果

論文の主張

- ▶ このような誤りをなくすために、「Statistically Significant（統計的に有意）は使わない」ようにしよう！（Wasserstain, 2019）。
- ▶ 次の4つの原則（ATOM）に基づいた解析と解析の評価を推奨する。

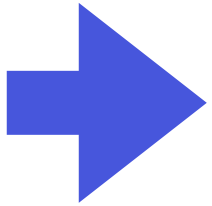
ATOMという原則 - I

- ▶ Accept Uncertainty（不確実性を受け入れる）
 - ▶ データの取得方法やモデルの仮定次第で、解析の結果は変動する。
 - ▶ 解析の結果にも、ばらつきがある（きちんと記載）。
 - ▶ 分位点 / 推定値の分散 / 推定値の信頼区間を必ず書く
- ▶ Be Thoughtful（思慮深く）
 - ▶ 【データの解析者が意識すべきこと】
 - ▶ 調査や解析の意図
 - ▶ 意味のある効果の大きさ
 - ▶ 目的に対する、データを取得方法の適切さ。
 - ▶ データに当てはめる手法と、手法の統計的性質の理解。
 - ▶ 例えば、線形回帰モデルでいえば、等分散性

ATOMという原則 - II

- ▶ Be Thoughtful（思慮深く） 続き
 - ▶ 【解析レポートを見る側が意識すべきこと】
 - ▶ 得られた推定値の、実際のな/実用的な意味
 - ▶ 推定値の正確さ（ばらつき）
 - ▶ 使用したモデルの仮定の適切性
 - ▶ 解析者のモデルに対する理解
 - ▶ 複数モデルの比較の有無 / 比較した場合の結果の変化と考察
 - ▶ 以上を、再現できるレベルで、レポートにまとめられているか

【ポイント】



データ解析の評価は、p値やその他の統計的尺度で行われるべきではない。
先行研究調査の深さ / 研究デザインとデータの質 / 仮定したメカニズムの
妥当性 / 現実的な価値 / 発見の新規性を総合して判断すべき

【アイデア】

結果をblindしてレポートを読んで価値を判断するのも1つの方法である

ATOMという原則 - III

- ▶ Be thoughtful（思慮深く） 続き2
 - ▶ モデルの妥当性評価については、p値以外にも提案されている
 - ▶ ベイズ因子 / 第2世代p値（Second Generation p-value）など
 - ▶ ただし、それは第2のp値を生み出すものであってはならない
- ▶ Be Open（公開する）
 - ▶ データ解析の価値は、データそのものの客観性。
 - ▶ しかし、データの取得から、解析に至るまで、解析者や専門家の「主観性」や「判断」に依存する部分が多い。
 - ▶ 結果を読み解くためには、背後にある「考え方」が必要で、それなしに第三者的なレビューが難しく、客観性が損われる。
 - ▶ 解析の客観性を保持するためには、データや解析のプロセスを公開し、置かれている想定などについて詳細なレポートを作成し報告するべき

ATOMという原則 - IV

- ▶ Be Modest（謙虚であろう）
 - ▶ 統計的な手法には、そもそも限界があることを理解する。
 - ▶ 統計モデルは複雑な現実を忠実に再現する手法ではなく、むしろ現象の「簡易模型」。
 - ▶ 使用されたモデルが「真のモデル」ではないことを理解しておく
 - ▶ 得られた結果は、「モデルが正しい！」という仮定のもとでしか意味のないもの
 - ▶ 自分の出した解析結果や、結論に対して「どんな間違いの可能性があるか」を考えておく必要がある。
- ▶ 研究においては、同じテーマに対して、複数の同様の研究が行われて、同じような結果が得られて、はじめて決定的なものとなる。
- ▶ 研究者は、再現性を保証するような研究を奨励すべきであり、その意味で研究の手順や、使用したデータについて、公表を行うべきである。

目次.

- ▶ イントロダクション
 - ▶ この発表で扱う内容と、発表のサマリー
- ▶ p 値 < 0.05 の罠。
 - ▶ p 値とは何か？
 - ▶ p 値を使う罠は「足元の想定」である（ここがポイント！）
 - ▶ p 値に対するASAのPrinciple
- ▶ p 値の問題を受けて、何をすべきか（Moving to a world beyond “ $p < 0.05$ ”）
 - ▶ 「統計的有意」からの脱却
 - ▶ ATOMという原則.
- ▶ 現実的なガイドラインとして（発表者の意見）
 - ▶ 私たちが、「 $p < 0.05$ の世界」から脱却するために

私たちはどうすればいいだろう？

- ▶ 自分がデータ解析しているトピックにおいて、「P値に価値があるかどうか？」という質問を考えることである。
- ▶ そのためには2つの質問に注目する。
 - ▶ データ取得は、きちんとデザインされており、RCT（ランダム化比較実験）であるか？
 - ▶ 解析に用いられるモデルの仮定は妥当あることが、先行文献などの結果から明らかであるか？
- ▶ いずれか一方がNOであるなら、P値を利用した仮説検定の結果から、安易に結論を導くのは適切とは言えない。
 - ▶ なぜなら、P値が「意味を持つ」のは、実験計画とモデルの仮定が妥当な場合であり、それ以外はP値を計算しただけである。

多くの解析は「目的」ではなく「データ」から始まる - I

- ▶ パターンは2つ。
 - ▶ (A) 目的があって、データの集め方からデザインする
 - ▶ (B) 手持ちのデータから、目的に対してアプローチする
- ▶ もちろん、(A) のほうがデータ解析としては適切なアプローチであるが、実際 (B) のようになってしまうのは、仕方のないこと！
 - ▶ 運動量が多い人ほど、疾病率が下がるか？という疑問に答える場合に、RCTまでやるというのはなかなか難しい。
 - ▶ 実際は、身近な人へ測定機器を配布して、歩行量と生体指標などを比較するしかない。
 - ▶ この場合の注意点は
 - ▶ 身近な人＝日本全国からのランダムサンプルではないので、まず結果は一般化できない。
 - ▶ そもそも測定機器を真面目に使うのは、健康意識の高い人なので、データ取得のバイアスが生じている。

多くの解析は「目的」ではなく「データ」から始まる - II

- ▶ サンプルの取得にバイアスがあることを踏まえた上で、関係性を見るためには、線形回帰モデルを当てはめて、回帰係数を見ればいい！

```
> summary(lm(y~x))
```

```
Call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-3.0332  -0.7744   0.0031   0.7196   5.7101
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.25552	0.03555	7.188	1.28e-12	***
x	0.58349	0.03527	16.543	< 2e-16	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.124 on 998 degrees of freedom
```

```
Multiple R-squared:  0.2152,    Adjusted R-squared:  0.2144
```

```
F-statistic: 273.7 on 1 and 998 DF,  p-value: < 2.2e-16
```

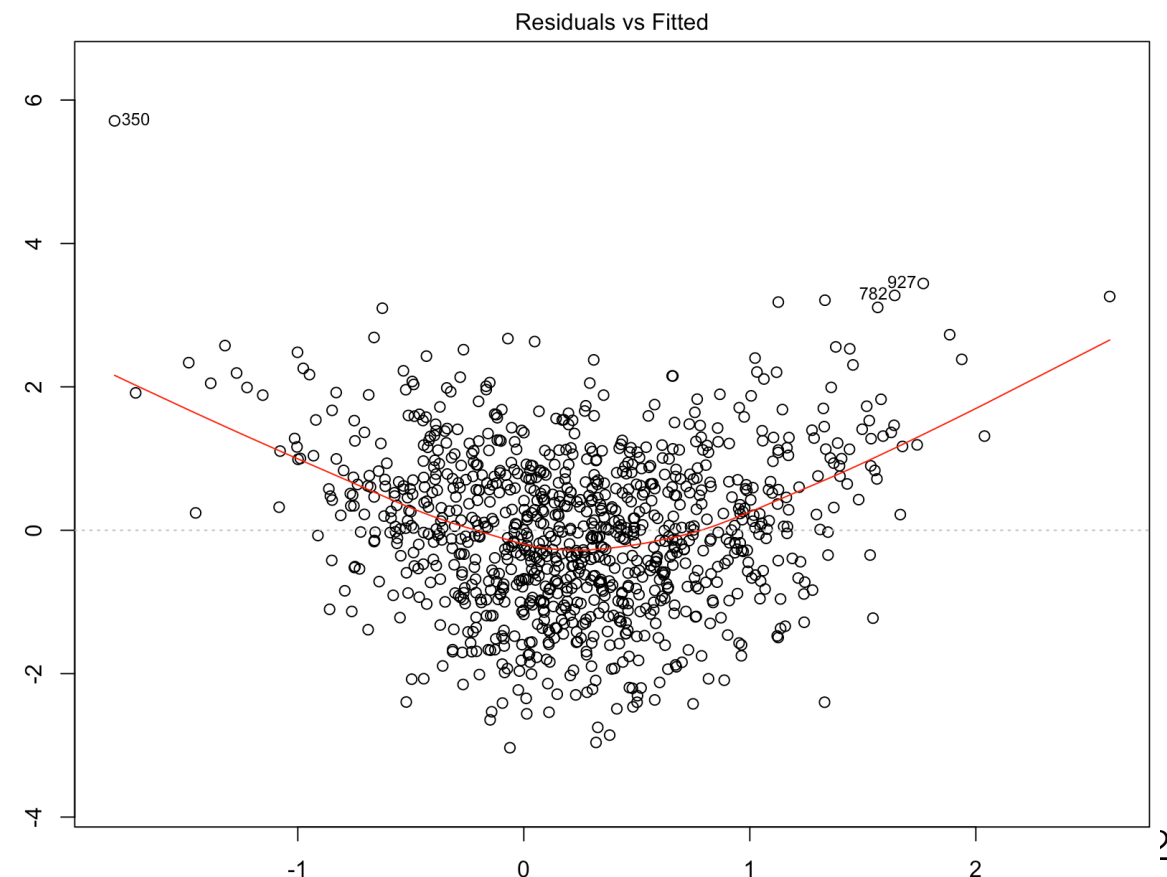
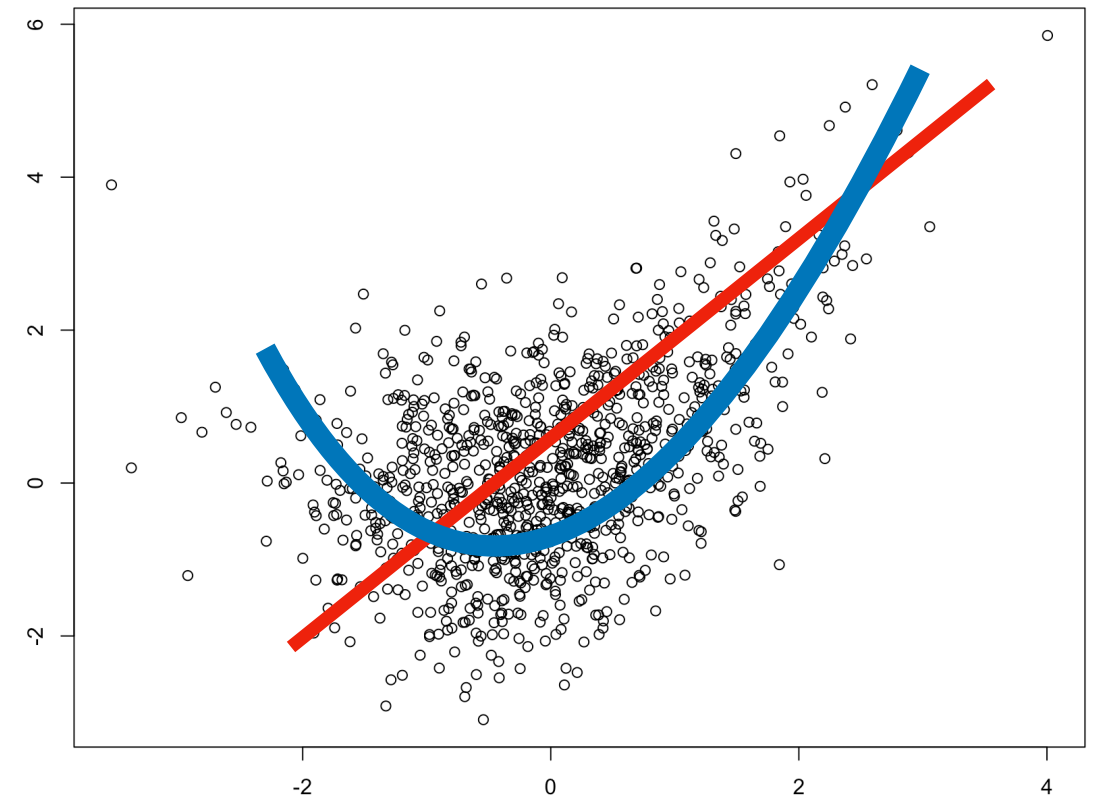
有意だ！！！！

ㄿ(๑òωó๑)ㄿ

- ▶ としてはいけなくて、「モデルの妥当性」をきちんと説明できるようにしていかなくてはならない。
 - ▶ そもそもXとYはどんな関係なんだっけ？
 - ▶ モデルの妥当性は？変数は本当に線形に効いてるの？などなど

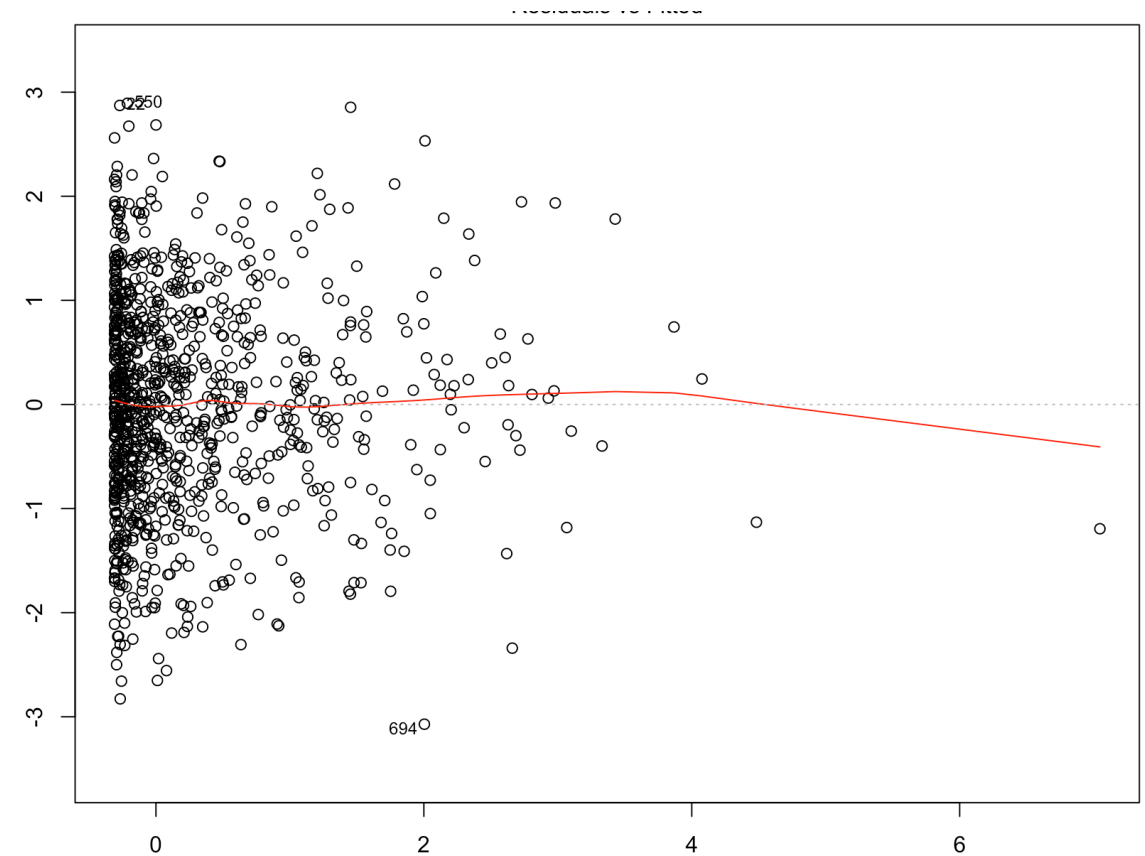
多くの解析は「目的」ではなく「データ」から始まる - III

- ▶ 先ほどの線形回帰モデルの結果は、右のデータを用いたもの。
- ▶ 確かに、何かの関係がありそうだけど、安易に線形にしているの？
- ▶ こういうときは、「線形回帰モデルを当てはめたときの仮定」を知っているかどうかで、判断ができる。
- ▶ 下の図は「当てはめ vs 残差」のプロット。線形回帰モデルの場合は「分散が一定」なので、傾向は残らないはず。
- ▶ しかし、明らかに2次の傾向が残っていて、これでは線形回帰モデルの仮定は満たされていない。



多くの解析は「目的」ではなく「データ」から始まる - IV

- ▶ 2次の項までモデルに含めると、残差に傾向が残らず、およそ分散は一定になった。
- ▶ 線形回帰モデルの仮定は、成立していると考えても良さそう。
- ▶ 次に、ここで推定された「回帰係数」が、「XからYへの影響」と考えられるとして、この回帰係数はどの程度ばらつくのか？
- ▶ この質問には次の2つを持って答える。
 - ▶ 1) 漸近分散と信頼区間の記載
 - ▶ 2) Bootstrap推定量のヒストグラムと、summaryを記載



```
> summary(lm(y~x+x2))
```

Call:

```
lm(formula = y ~ x + x2)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.07137	-0.69371	-0.00291	0.71458	2.89301

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.05367	0.03904	-1.375	0.17
x	0.55704	0.03223	17.285	<2e-16 ***
x2	0.30364	0.02135	14.224	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.025 on 997 degrees of freedom

Multiple R-squared: 0.3476, Adjusted R-squared: 0.3463

F-statistic: 265.6 on 2 and 997 DF, p-value: < 2.2e-16

多くの解析は「目的」ではなく「データ」から始まる - V

- ▶ データ解析の終わりには、次の2つは必ずチェックする。
 - ▶ 「統計的な仮定」に対する矛盾は最小限に抑えられているか？
 - ▶ データ解析の結果と、現実的な私たちの感覚に大きすぎる齟齬はないか？
- ▶ レポートの作成
 - ▶ まず、データの取得と、今回の解析結果の適用可能範囲を明記する。
 - ▶ また、どのような解析を検討し、途中でどんな結果を得て、最終結果に至ったかを明記する。
 - ▶ 現実的な感覚との齟齬がないか、齟齬があるのであればどんな原因かを書く。
 - ▶ 例えば、「〇〇のような変数が足りてない」など

データの解析を行ったレポートで、推定量のばらつきの記載がない場合、それはデータ解析ではない。

まとめ

- ▶ ここまで、データ解析の一例を示してきたが、特に（B）**手持ちのデータから、目的に対してアプローチする**データ解析では、
 - ▶ 解析対象となる現象に対して、深い知識が必要
 - ▶ 統計学に対する深い理解と、その応用能力が必要。
- ▶ これら2つで、データ解析である。まずは、自分たちの行っている「データ解析」のフローを見直してほしい。

チェック項目

- ▶ P値や他の統計的な基準を、主張の根拠としていないか。
- ▶ 使用している解析ツールへ十分な理解があり、正しい運用を行っているか（また、解析後残差プロットなどをチェックし、レポートへ掲載しているか？）。
- ▶ 取得されたデータにバイアスは存在していないか？
- ▶ 結果の過剰な一般化は、行われていないか？
- ▶ 解析結果に、漸近分散/信頼区間/Bootstrap推定量のばらつきのプロットの3つはきちんと掲載しているか？

Let's move to a world beyond “ $p < 0.05$ ” together!!

ありがとうございました！

慶應義塾大学大学院

中村 知繁

tomoshige.nakamura@gmail.com

参考文献

- ▶ Amrhein, V., Greenland, S., and McShane, B.(2019).Scientists rise up against statistical significance. Nature, 567, 305-307.
- ▶ Fisher, A.(1935). Statistical Test, Nature 136, 474.
- ▶ Held, L., and Ott, M.(2018). On p-Values and Bayes Factors. Annu. Rev. Stat. Appl., 5, 393-419.
- ▶ Hung, J., O'Neill, R., Bauer, P., and Kohne, K. (1997). The Behavior of the P-Value When the Alternative Hypothesis is True. Biometrics, 53, 11-22.
- ▶ Sellke, T., Bayarri, S., and Berger, J. (2001). Calibration of p Values for Testing Precise Null Hypothesis. The American Statistician, 55, 62-71.
- ▶ Wasserstein, R., and Lazar, N. (2016), The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70, 129-133.
- ▶ Wasserstein, R., Schirm, A., and Lazar, N. (2019) Moving to a World Beyond " $p < 0.05$ ", The American Statistician, 73, 1-19.
- ▶ 柳川 堯. (2017). p値は臨床研究データ解析結果報告に有用な優れたモノサシである. 計量生物学, 38, 153-161.