

※スライドの一部だけをWebなどで引用する（画像を貼るなど）のはご遠慮ください

再現可能性にかかわる 統計的問題と 新しい論文のカタチ

清水裕士

関西学院大学社会学部

第84回日本心理学会シンポジウム 2020.11.1

再現可能な心理学研究入門

自己紹介

- 清水裕士
 - 関西学院大学社会学部
- 専門
 - 社会心理学
- Web
 - <https://norimune.net>
 - Twitter:@simizu706
 - HADという統計ソフトを作っています
 - <https://norimune.net/had>

今日話すこと

- 帰無仮説検定に論理と信用の問題
 - 論理の問題
 - 信用の問題
- ポスト帰無仮説検定の世界
 - どんな論文のカタチがありえるだろう？

「再現可能」な統計分析？

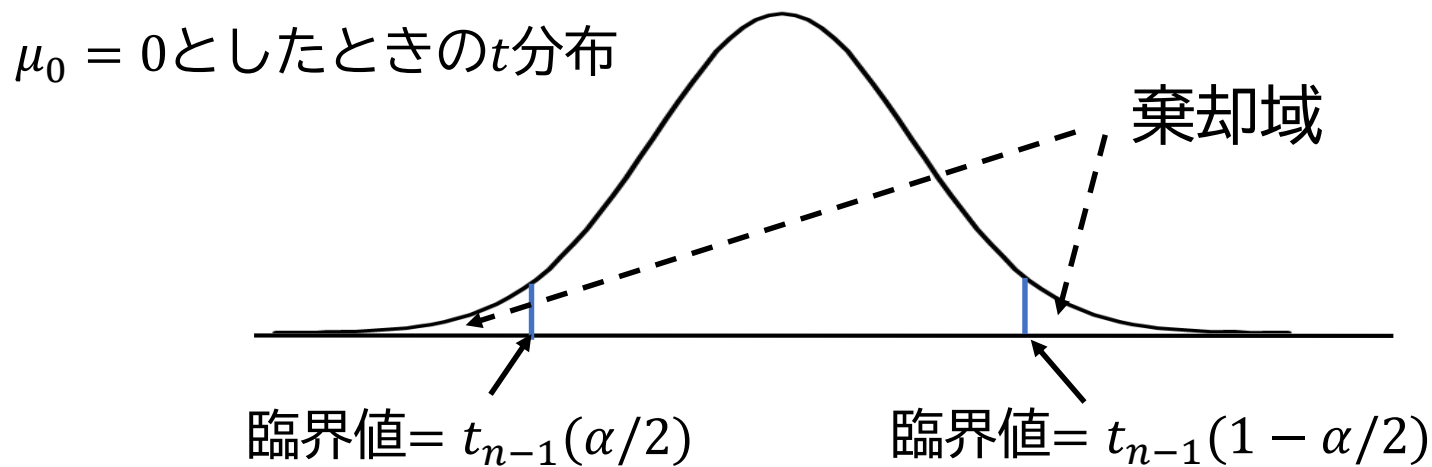
- 「再現可能性」という言葉はキレ味がよすぎる
 - 心理学者にかなりの「温度差」を生む言葉
 - ハードサイエンスを強く志向する立場か否かでも
 - そのわりに再現可能性の定義がイマイチわからない
- 「信頼できる」統計分析
 - 統計学の論理は、ある程度はっきり白黒つけられる
 - しかし、「再現可能性」の議論は統計の論理を超えた、信用の問題に関わる
- 信頼できる = 論理的に正しい + 信用できる

帰無仮説有意性検定(NHST)

- 帰納推論を統計的に行う
 - 帰納推論は論理的には正当性を持たない
 - それを確率の概念を用いて正当化を行う体系
- ある確率モデルのもと、統計的過誤確率 α, β を定めることで、推論の誤りをコントロールする
- 正当化のための基準
 - 検定の論理
 - 検定の手続き自体の正しさの問題
 - 検定の信用性
 - 検定の手続きが正しく行われているかどうかの判断の問題

検定における「論理の問題」

- 「事前」に決めておくべきもの
 - 確率モデル（正規分布→帰無分布が t 分布）
 - 帰無仮説（ $\mu_0 = 0$ ）
 - サンプルサイズ(n)
 - 有意水準(α)



検定の論理としてアウトなこと

- サンプルサイズの事後変更
 - 有意になるまでサンプルサイズを増やす
 - 一部の参加者だけを使う
 - *p* hacking
- 帰無仮説の事後変更
 - 帰無仮説を分析結果を見て変える
 - 複数行われた検定の一部だけを報告する
 - HARKing
- 正しい分析をしよう

サンプルサイズ設計

- サンプルサイズ設計とは
 - 有意水準と検出力、検出する効果量がわかっている場合、必要なサンプルサイズがわかる
 - 大きすぎるサンプルをとる必要がなくなる
- 検出力の確保と、必要最低限のサンプル
 - 検出力（有意差が出せる力）は80%以上が目安
 - ある効果量を出すためのサンプルサイズがたとえば100だった場合、それ以上取らなくていいと判断
 - 無駄にリソースを消費しなくていい

持ちかけられる相談(仮想)

- サンプルサイズ大きすぎ問題
 - サンプルサイズ設計ちゃんと計算したら $n=235$ になりました
 - 無理です
 - サンプルサイズ設計はやめておいて、各群40で行こうと思うのですがいいですか？
- 検出力計算できない問題
 - GLMMの検定力を計算するパッケージがない
 - 無理です
 - サンプルサイズ設計はやめておいて (以下略)

正直に書こう

- サンプルサイズは「事前に決める」ことが大事
 - 検出力が80%ないと検定が誤りなわけではない
 - 厳密なNP流の仮説検定では求められるが、現代的な帰無仮説検定では帰無仮説を採択しないので、必須ではない
 - どれくらいのサイズをとれるかがわかってるなら、そこから検出力がいくらになるのかは計算可能
 - それらを正直に報告すればいい
- 事前に決めたサイズがとれなかった
 - まずはがんばってとろう
 - なぜとれなかったのか正直に書こう

事前の計画の問題

- 帰無仮説検定は「事前」にあれこれ決めておく
 - どの群とどの群の差を比較するのか
 - 帰無仮説
 - 確率分布はどうするか、共変量がある場合はどれか
 - 確率モデル
- なぜ事前決めないといけないのか
 - 検定結果を見てから計画を変えると、帰無分布が理論通りの分布（たとえば t 分布）に従わなくなる
 - →第一種の過誤のインフレの原因となる

「事前」って何の「前」？

- データ取得の前？
 - 実験の前に計画を立てるのは重要
 - しかし2次データ分析の場合は？
- 本質的には「帰無分布が変わらない」が重要
 - 実験した後でも、ローデータを見たとしても、帰無分布が変わらなければ問題はない
 - つまり、データの情報を得ることで検定の計画が変わってなければ、検定は「論理的」には正しい結果を返す

正直に書こう

- 計画以外の変数もはいつてるんですけど
 - 共同研究で、いろんな変数を測定している
 - すべての変数について α を調整しないといけない？
 - 何を同時に測定したのか、事前の計画はなんだったのか正直に書こう
- 結果見てからいいこと思いついたんですけど
 - 検定結果を見てから、あたらしい分析を思いついたなら、そのように正直に書いて分析をしよう
 - ただ、事後計画の検定は α の調整が必要
 - 調整がわからなければ、それについては検定をしないか、結論を保留しよう
 - 次に新しくとったデータで検証する

検定における「信用の問題」

- 「データ見たけど計画は変えてないで」
 - という著者を我々は信用できるか
 - おそらく再現可能性問題以後、我々はそれを信用しない（できない）ことにしたのだろう
 - ズルするやつがいることがわかったから
- 正しく、正直に書くことだけではダメなのか
 - 後知恵バイアス
 - 嘘をつくインセンティブ

研究者の自由度

- その研究者しか知りえない情報と操作可能性
 - 事前計画、データ収集手続き、データそのもの
 - これらを自由に操作できると、研究者にとって不当に有利な論文執筆が可能になる
- 研究者の自由度が高いと、信用できない？
 - 研究者はズルをするかもしれないから、自由に操作できる研究計画は信用しない方がいい
 - 正直に書いた論文に「お前は事後的に計画を変えただろう」と査読者に言われたらどうしようか？

自由度を下げるために

- オープンソース・オープンデータ
 - マテリアルやデータそのものをオープンにする
 - これも研究者の自由度を下げることに寄与する
 - もちろんほかの人が使えるようになるというメリットも
- 事前登録
 - データをとる前に計画を書いて登録
 - 検定における「研究者の自由度」を大幅に下げることができる
- 帰無仮説検定をやめる
 - 実は、結構本質的な解決策ではある
 - 後述

「信頼できる分析」結果の書き方

- 正しく
 - 検定の論理を守ろう
- 正直に
 - やったことを正直に書こう
- オープンに
 - できるだけオープンにして共有しよう
- 事前登録
 - できれば事前登録をして信用してもらおう

信用の問題は難しい

- 完全に信用させるのは無理がある
 - どんな厳しい制度もhackingは可能
 - 信用基準を厳しくしすぎると、正しく・正直に書くインセンティブが低下する
 - 「これさえやっとけばいい」になるのも問題
- コミュニティで考える必要
 - 検出力をどの程度にするのか
 - オープンデータ・事前登録は必須なのか
- 信用基準のレベルで分断が生じることは心理学においてあまりいいことにはならないのでは？

まとめ

- 帰無仮説検定の論理と信用の問題
 - 正しく
 - 正直に
 - オープンに
 - 事前登録

この発表に関する資料

- 土居淳子（2010）
 - [帰納的推論ツールとしての統計的仮説検定：有意性検定論争と統計改革](#)
- Goodman, Fanelli, & Loannidis(2016)
 - [What does research reproducibility mean?](#)
- 南風原朝和（2014）
 - [続・心理統計学の基礎 総合的理解を広げ深める](#)
- 岡田謙介先生のスライド
 - [「 \$p < 0.05\$ 」からATOMへ 不確実性を受容する統計的方法](#)
- 大塚淳（2020）
 - [統計学を哲学する](#)