疑わしい研究実践に陥らないために
ー心理学の経験からー

慶應義塾大学文学部
平石 界
Special thanks to 再現性科研メンバー
（三浦麻子・樋口匡貴・藤島喜嗣）

本日の内容

- 動画の背景
  - 心理学における再現性危機
  - QRPs/P-hackingの“発見”
- 動画で十分に触れられなかった点
  - HARKingの何が問題か
  - 事前登録、レジレポ
- 心理学分野の教訓
  - どんなシステムもハックされる
  - 風通しの良さとマルチレイヤー指導

Reproducibility Project: Psychology

- 2008年に主要雑誌に掲載された論文を追試（Open Science Collaboration, 2015, *Science*）[1]。
- 「追試成功率」＜40%
- オリジナル研究と比べて、効果量もことごとく低い。
- **主要誌の研究の半分が再現されないんだったら、何を信じれば良いのか。**

Science WEBサイト:

https://www.science.org/doi/10.1126/science.aac4716

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). , Figure 3

Do you believe in psi?

- 超能力論文（2011年）
  - Bem, D. J. (2011). “Feeling the Future” *Journal of Personality and Social Psychology*[2]
  - 著者はCornellの教授。社会心理学の有名研究者。
  - 掲載誌は社会心理学領域のトップジャーナル。
  - セクシーな画像がどっちに出るか予測できる！
  - 未来に学ぶことは、その前から記憶されている！9つの実験を行って、そのうち8つで統計的に有意（$p < .05$）な結果
- 追試（失敗）報告をJPSPがdesk rejectしたことが混乱に拍車を掛けた[3]。
  - *JPSP*は当初からEditorial Comment[4]と批判論文[5,6]もセットで公刊したこと、後日、別の追試論文（タイトルは“Correcting the Past”[7]）を掲載したことは書いておきます。
- 一大追試ムーブへ（前のスライドに続く）

AIP WEBサイト:

http://doi.org/10.1063/1.3663724

Bem, D. J. (2011). Feeling the Future: Experimental evidence for anomalous retroactive influences on human cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. Table 7

音楽で若返り

- 大学生20人に“Kalimba”または“When I‘m Sixty Four”を聴かせた。
- それから参加者の生年月日と、参加者の父親の年齢を尋ねた。
- 父親の年齢を統制して分析したところ（ANCOVA）、“When…”を聴いた条件の方が、年齢が、約1.5才若くなっていた（p = .04）。
- Simmons, Nelson, & Simonsohn (2012) “False Positive Psychology” [8]
- そんなバカな！？

論文著者の
Joseph Simmons

論文著者の
Leif D. Nelson

When I‘m Sixty Four

論文著者の
Uri Simonsohn

真実はいつもひとつ…！？

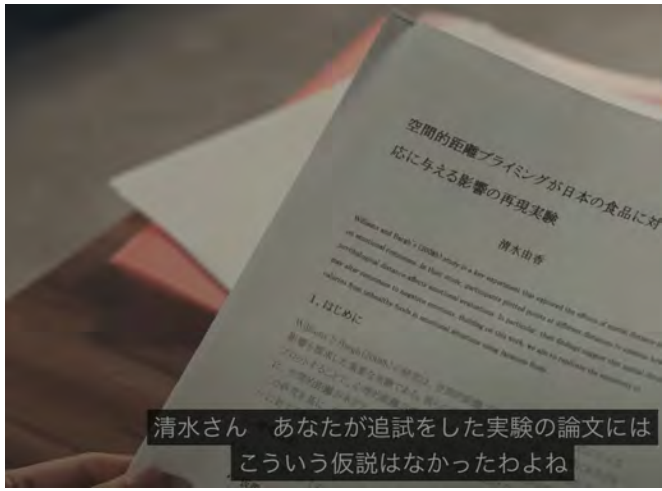
- ペンシルバニア大の~~34人~~20人の学生に、“Hot potato”, “When I‘m sixty four”または“Kalimba”を聴かせた。参加者が10人集まるごとに統計をかけた。事前にサンプルサイズは決めていなかった。それから別の調査と言って本人の生年月日、最近年を取ったと感じる程度、ダイナーでの食事が好きな程度、100の平方根、“パソコンは複雑な機械だ”と思う程度、父親の年齢、母親の年齢、早朝割引を使うか程度、政治的指向、カナダ人の4人のクオーターバックの誰が賞を得ると思うか、昔のことを「古き好き日々」と思う程度、そして性別を尋ねた。父親の年齢を統制したところ・・・（以下略）。

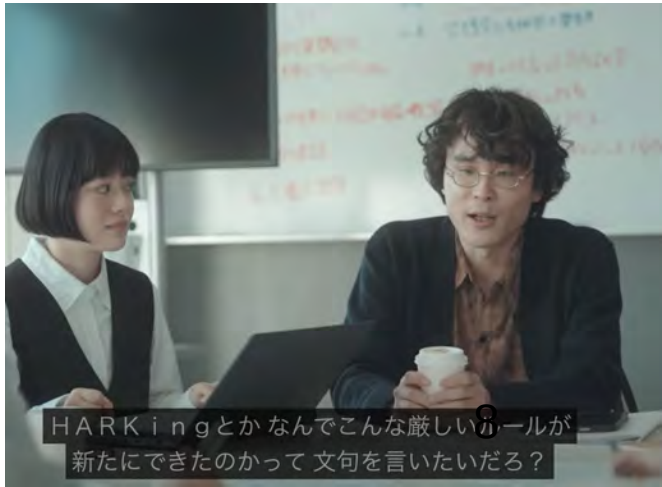
真実はいつもひとつ…！？

- **ペンシルバニア大の~~34人~~20人の学生に、“Hot potato”, “When I‘m sixty four”または“Kalimba”を聴かせた。参加者が10人集まるごとに統計をかけた。事前にサンプルサイズは決めていなかった。それから別の調査と言って本人の生年月日、最近年を取ったと感じる程度、ダイナーでの食事が好きな程度、100の平方根、“パソコンは複雑な機械だ”と思う程度、父親の年齢、母親の年齢、早朝割引を使うか程度、政治的指向、カナダ人の4人のクオーターバックの誰が賞を得ると思うか、昔のことを「古き好き日々」と思う程度、そして性別を尋ねた。父親の年齢を統制したところ・・・（以下略）。**

Questionable Research Practices

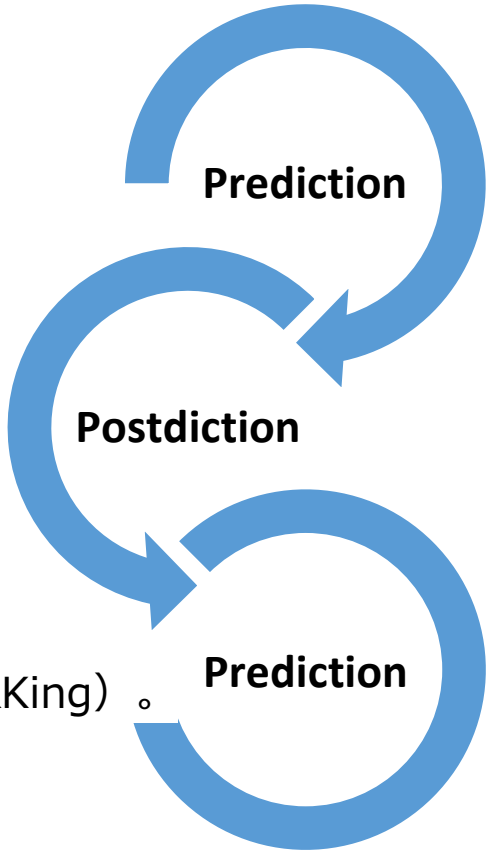
- **Multiple testing**
  - 手当たり次第に分析をかけ、有意になった組み合わせだけ報告。
- Underpowered study
  - サンプルサイズの小さい実験をいくつも繰り返し有意になったものだけ報告。
- Optional stopping
  - データを少しずつ足しながら、都度検定をかける。有意になったら終了して報告する。
- これらを駆使すれば、ほぼ確実にどこかで統計的に有意な結果を得ることができる。
  - ただしそれらはただの偶然（偽陽性）である可能性が有意水準よりも高い。
- **HARKing**
  - 予期せぬ分析結果を“解釈”し「最初からそう予測をしていた」という報告する。
- しかし、それこそが「データを深く読み込んだ、優れた研究」と思われてきた。





PredictionとPostdiction

- **Pre**diction
  - 理論や仮説から導かれた、データを見る前の予測。
  - 仮説検証研究（confirmatory research）で用いる。
  - 既存の仮説や理論の正しさを検証する。
- **Post**diction
  - 既存データから生まれる、説明や解釈。
  - 探索型研究（exploratory research）から得られる。
  - 新たな仮説や理論を生成する。
- **Pre**dictionと**Post**dictionを繰り返すことで科学は進歩する。
  - Predictionだけでは、新発見は生まれない。
  - Postdictionだけでは、言ったもの勝ちになってしまう。
- **Pre**dictionと**Post**dictionは混同されやすい。
  - しかし混同してしまうと「言ったもの勝ち」になってしまう（HARKing）。
- 混同を防止するのが事前登録システム。
  - データを見る前にPredictionを登録してしまう。
  - データを見てから思いついたことはPostdictionであることが明確になる。
- Nosek, et al. (2018)[9]



Ulysses and the Sirens by H.J. Draper[10]

事前登録
Pre-registration

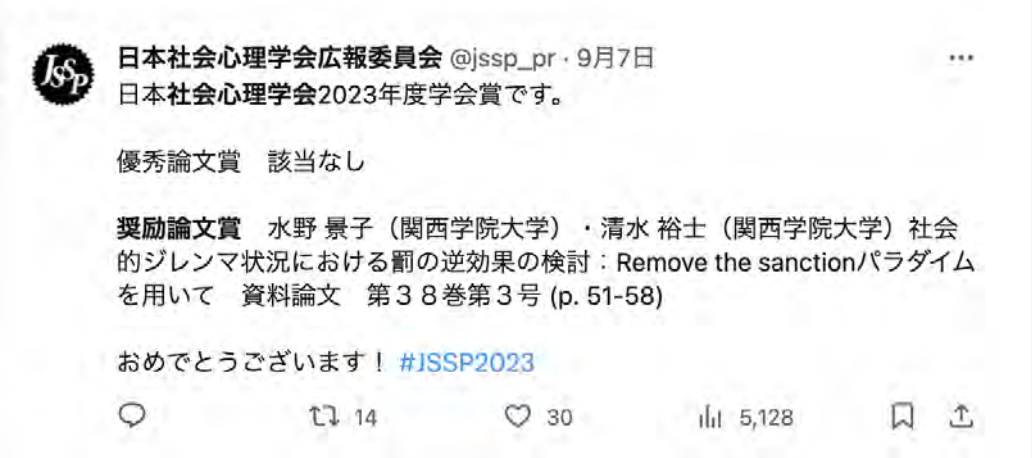
- データ取得前／分析前に、研究計画をタイムスタンプ付きで登録してしまう
  - 研究の種類、実験デザイン
  - 独立変数、従属変数の測定方法
  - 分析手法、共変量、多重検定の調整法
  - サンプルのリクルート方法、サンプルサイズ、除外基準など
- 事前登録した内容は、研究者本人であっても改変不可能（改変すると記録が残る）にする。
- つまり、自分で自分のやること（データ取得、分析、解釈プロセス）に“縛り”をかける。
- 縛りを抜けた場合に、それが自他に明らかになるのがポイント。

でも…

- どこに登録したらよいか分からないよ。
  - Open Science Framework (https://osf.io/)
  - As.Predicted (aspredicted.org)
- 何を書いたらよいか良いのかわからないよ。
  - 長谷川ほか (2021) 心理学研究[11]
  - Bosnjak et al. (2021)[12]の阪大社心教室有志の邦訳版[13]
  - 計画どおり行かないこともあるじゃない。
  - 修正のあったことも含めて報告すればよい
  - 「ありのままを伝えよ」（樋口・三浦・藤島による邦訳）[14]
- どう決めたら良いのかわからないよ。
  - 探索段階であることを認めるのが吉。
  - 電子ラボノートを使うのも一案だろう。
- アイディアが盗まれたら…
  - 事前登録の公開時期を遅らせる（embargo）が可能。
- 二次データ分析なのですが。
  - どこまでデータを見たうえでの分析なのか宣言するのがよい（PCI-RRのポリシー）[15]。

それなら事前登録から指導しておきなさいって話
※ストーリー展開などの関係上、敢えて教員にも隙を残しました。

JSTAGE WEBサイト:
https://www.jstage.jst.go.jp/article/jssp/38/3/38_2205/_article/-char/ja/
社会的ジレンマ状況における罰の逆効果の検討：
Remove the sanctionパラダイムを用いて
水野 景子, 清水 裕士



社会心理学会ではプレレジ付きの追試論文（結果はネガティブ）が授賞

- 非協力者への罰の逆効果を徹底的に厳密に追試したもの。
- その徹底的な厳密さがあるからこその論文掲載であり、授賞であることは、強調しておきます。
- 水野景子, & 清水裕士. (2023). 社会心理学研究, *38*(3), 51–58.[16]
- https://osf.io/k9pf4

事前登録があれば万全？

- **P**reregistration **A**fter the **R**esults are **K**nown
  - 実験をやって結果が分かってから事前登録をすることで「予測通りの結果が出ました」と報告する。
  - Yamada, Y. (2018). [17]
- 大量事前登録という手段もある。
  - 実施予定の実験を片端から事前登録しておいて、有意な結果が得られたものだけを“事前登録付き”で報告する（事前登録付き実験リセマラ）。
- 統計的有意性へのバイアス補正機構が組み込まれてない。
  - 「事前登録して実験したら有意じゃなかったです」
  - 「ふうん」「で？」実験がまずかっただけでしょ。だって～が共変量に入ってない」「リジェクトね」

Registered Report

- 研究計画を審査する（Stage 1）。
  - 査読を通った計画を In Principle Accept（IPA, 原則受理）する。
  - 結果が有意でも、有意でなくても、雑誌に掲載されることを約束する。
- 研究結果を含む原稿を審査する（Stage 2）。
  - 計画通りに研究が実施されたかを審査する。
- 審美的基準による雑誌への掲載判定を食い止めることができる。
  - 研究計画への評価の段階で、審美的基準が入り込むことは止められない（「本当だったら驚き」な研究計画ほど、高く評価されるバイアスは残る）。

Nature WEBサイト:
https://www.nature.com/articles/s41562-021-01193-7

Christopher D. Chambers et al,
The past, present and future of Registered Reports

Nature Human Behaviour volume 6, pages29–42 (2022)
Figure 1a

Christopher D. Chambers et al,
https://www.nature.com/articles/s41562-021-01193-7 [18]

マルチレイヤーの指導

- どんなシステムもhackされる。
  - 事前登録もhackされうる。
  - ピアレビューが仲間内レビューによって骨抜きにされているとの批判もある。
  - registered reportも、いずれhackされる（hackしてしまう）だろう。
  - 研究者という仕事の意味をしみじみ考える時間を持つ必要がある（そういう時間をシステムとして組み込む必要がある）。
- 明立大学心理学部の良かった点
  - 複数スタッフが、それぞれに異なる立場とアプローチから学生をサポートをしていたこと。
  - 複数スタッフが互いを信頼して、自らの主張をし、他者の主張を受け止めていたこと。

事前にいただいたご質問について

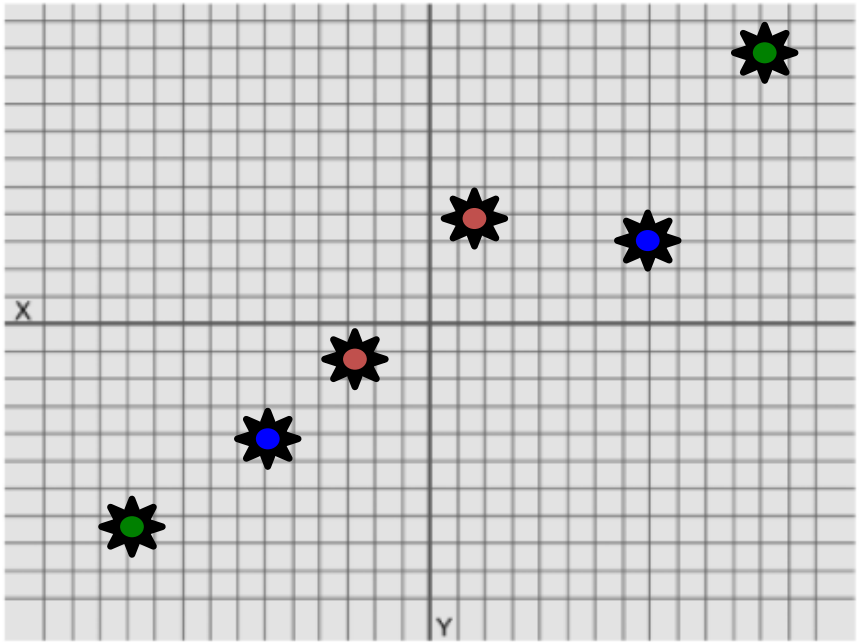
- グループディスカッションのポイント
  - QRPの何がまずいかの再確認でしょうか。
  - HARKingを防ぐ方法を考えてもらうのもよいかも。
  - 事務の方がテスト作成をする体制が問題ありとも思えます（逆を考えれば明らか）。
- 学生教育について
  - 方法論テキストに章を設けたりもしていますが…。
- 人文学（哲学など）の研究倫理について
  - 専門外なので迂闊なことは言えません。
  - 直接の研究対象者でない人々（個人・集団）への影響まで考慮に含めるべきという提言は出ています（NHB, 2022）[19]。
  - 資料研究などで個人情報に触れてしまうリスクに備えることも、研究倫理として求められていくかも知れません。

以下はおまけ

動画の元ネタとなった藤島・樋口（2016）論文について

真実の＊を探す旅

- 距離プライミング
  - 近接条件：(2,4)と(-3, -1)
  - 中庸条件：(8,3) と(−6,—5)
  - 遠隔条件：(12,10)と(−11,−8)
- 健康食品のカロリー評定
  - 近接プライム＝遠方プライム
- 不健康食品のカロリー評定
  - 近接プライム＞遠方プライム
- Williams, L. E., & Bargh, J. A. (2008). Keeping one’s distance: the influence of spatial distance cues on affect and evaluation. *Psychological Science, 19(3), 302–308.* [20]



Sage Journals WEBサイト:
https://doi.org/10.1111/j.1467-9280.2008.02084.x
Williams, L. E., & Bargh, J. A. (2008). Keeping one’s distance: the influence of spatial distance cues on affect and evaluation. *Psychological Science*, *19*(3), 302–308.
Figure 3

追試してみた

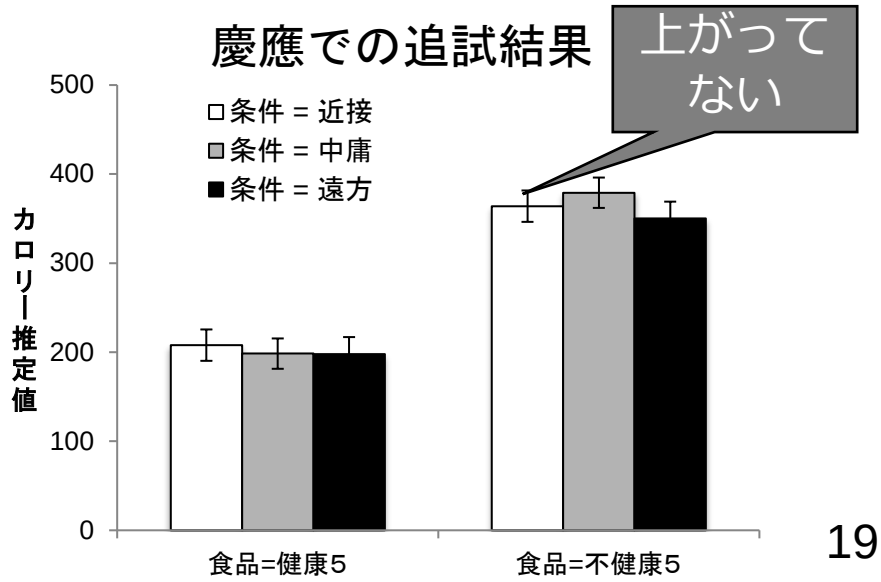
- 元研究より大きな検定力で実施。
  - 元研究：n = 59
  - 追試：４大学合計で n = 396
  - 結果は再現されず。
  - 元研究と同じ分析（交互作用）は有意にならず（$p > .05$）。
- **「でも、何か今回に特殊な事情があったのかも知れない！」**
- 藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における“*p*-hacking”の実践例. 心理学評論, *59*(1), 84–97. [21]

https://www.jstage.jst.go.jp/article/sjpr/59/1/59_84/_pdf/-char/ja
藤島喜嗣, & 樋口匡貴. (2016).
社会心理学における“*p*-hacking”の実践例. 心理学評論, *59*(1), 84–97.
Table 3をグラフ化

白が黒より上がっているのがポイント

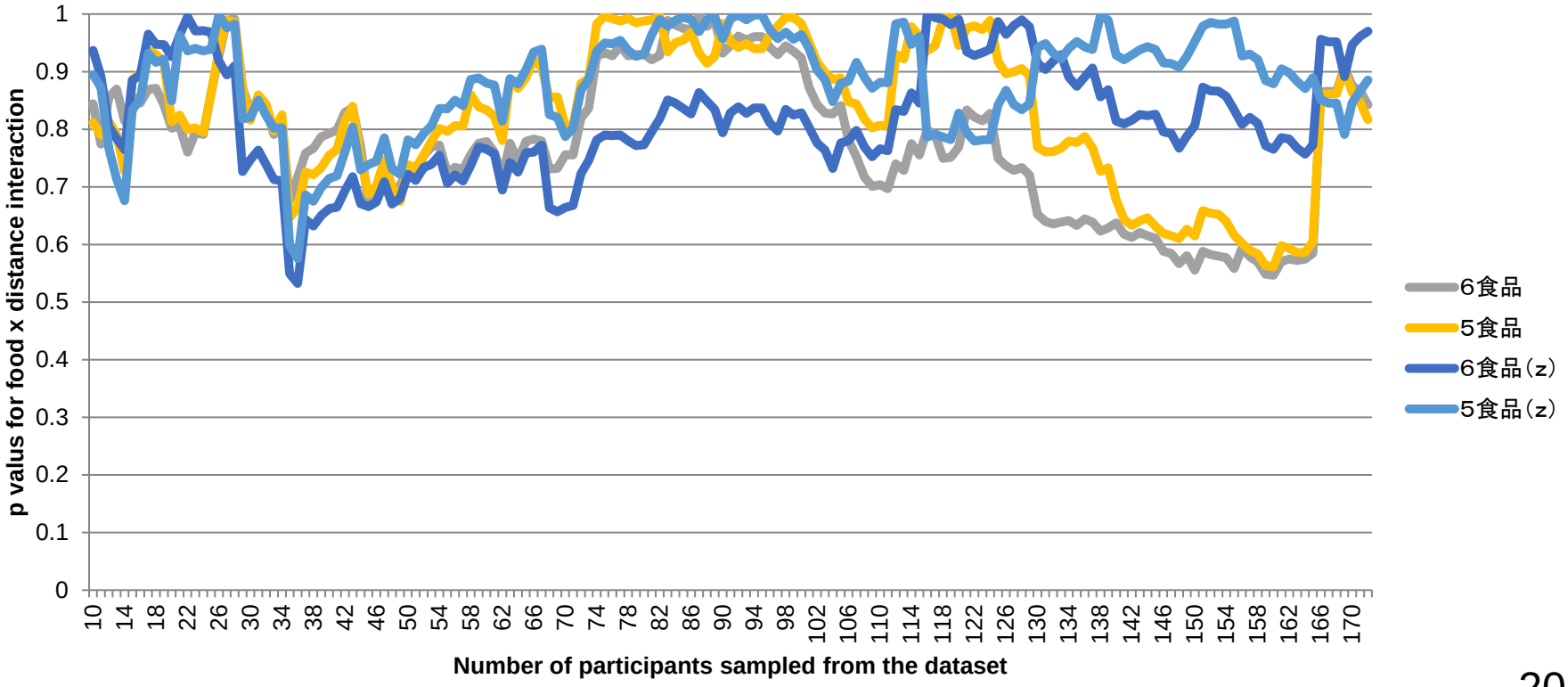
Sage Journals WEBサイト:

https://doi.org/10.1111/j.1467-9280.2008.02084.x
Williams, L. E., & Bargh, J. A. (2008). Keeping one’s distance: the influence of spatial distance cues on affect and evaluation. *Psychological Science*, *19*(3), 302–308.
Figure 3

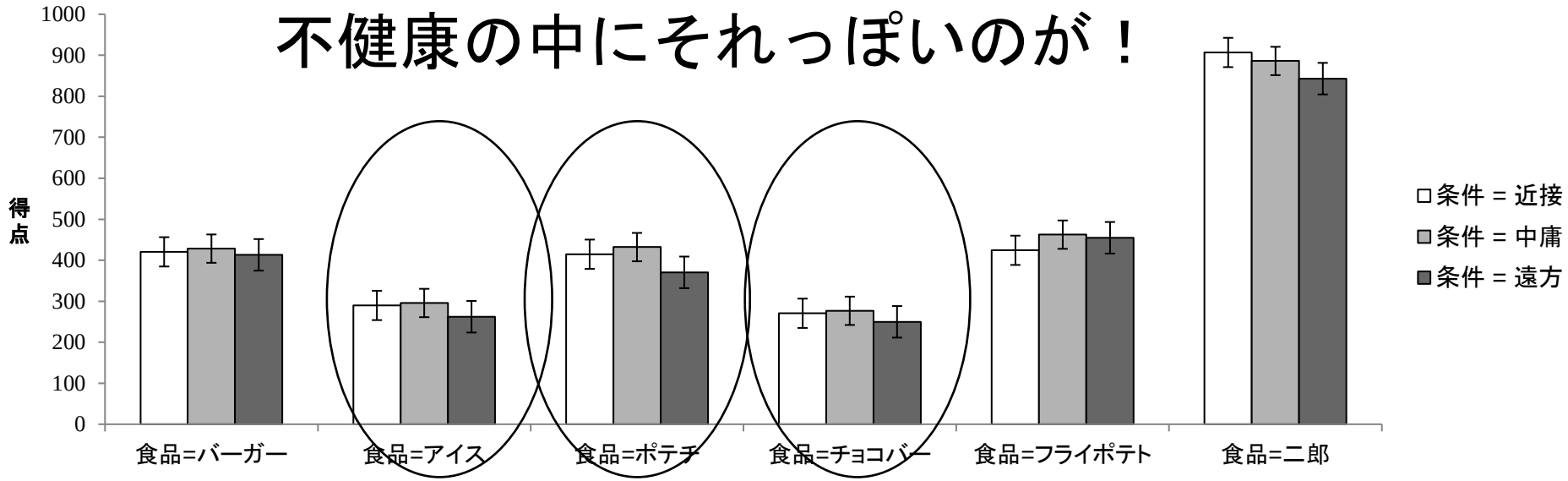


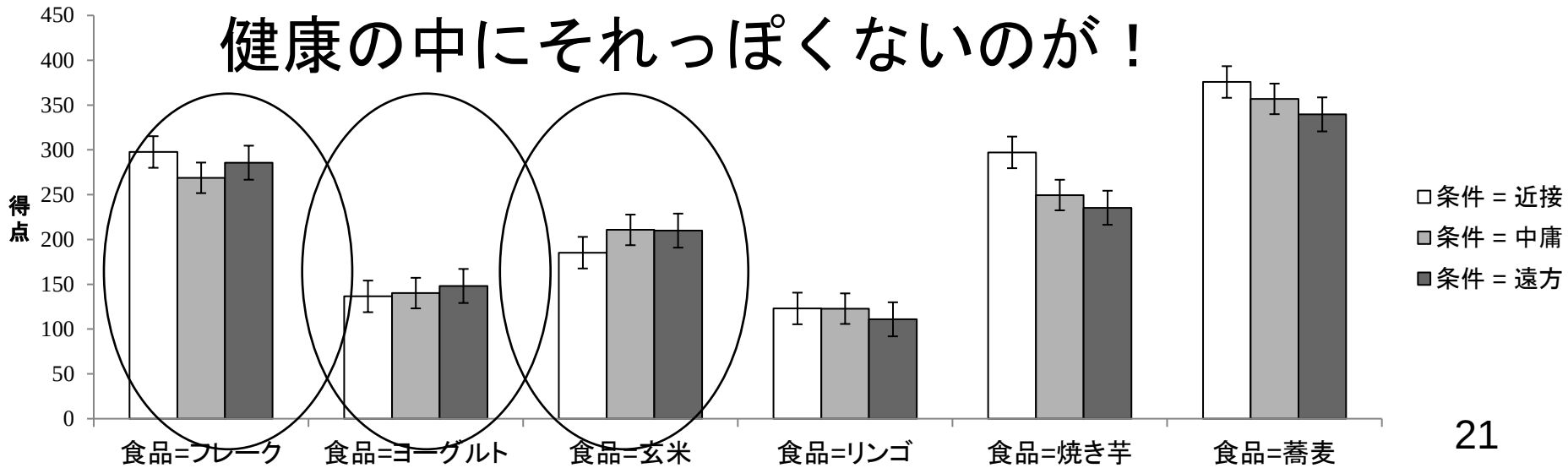
Sequential Analysisやってみた

- データを取った順に少しずつ使って、その都度、ANOVAやってみればいいじゃない！(^^)
- そうは問屋がおろさない。



データの長所を探してみよう！



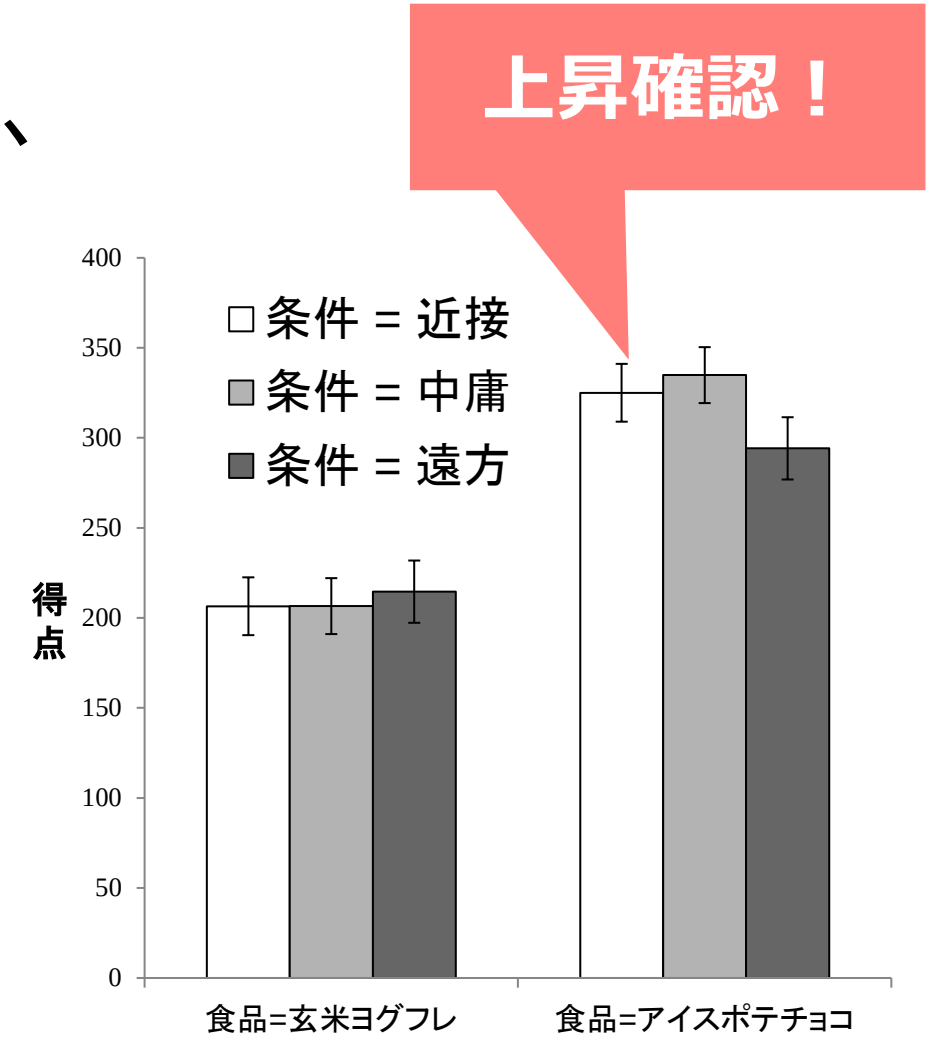


https://www.jstage.jst.go.jp/article/sjpr/59/1/59_84/_pdf/-char/ja

藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における“*p*-hacking”の実践例. 心理学評論, *59*(1), 84–97. Table 2をグラフ化

HARKingに踊れば、

- 不健康食品３つ
  - アイス、ポテチ、チョコバー
  - 間食系だ！
- 健康食品３つ
  - 玄米、ヨーグルト、コーンフレーク
  - 主食系だ！
- 言われてみれば！



交互作用　p = .061

https://www.jstage.jst.go.jp/article/sjpr/59/1/59_84/_pdf/-char/ja
藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における“*p*-hacking”の実践例. 心理学評論, *59*(1), 84–97.
Table ２をグラフ化

樋口・藤島(2016)[22]

※悪い冗談です（為念）。

JSTAGE WEBサイト:

https://doi.org/10.24602/sjpr.59.1_84

藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における“*p*-hacking”の実践例. 心理学評論, *59*(1), 84–97.
p.87

References

1.Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

2. Bem, D. J. (2011). Feeling the Future: Experimental evidence for anomalous retroactive influences on human cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. http://doi.org/10.1063/1.3663724

3. Chris French (2012).Precognition studies and the curse of the failed replications. The Gardian, https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications

4. APA PsycArticles: Journal Article, Editorial comment. https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022789

5. Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Kievit, R., & van der Maas, H. (2018). Yes, psychologists must change the way they analyze their data: Clarifications for Bem, Utts, and Johnson (2011). In *PsyArXiv*. https://doi.org/10.31234/osf.io/tvarg

6. Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

7. Galak, J., Leboeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the Past: Failures to Replicate Psi. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/a0029709

8. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

9. Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (11): 2600–2606. https://doi.org/10.1073/pnas.1708274114.

10. Ulysses and the Sirens by H.J. Draper
https://commons.wikimedia.org/wiki/File:Ulysses_and_the_Sirens_by_H.J._Draper.jpg
Herbert James Draper, Public domain, via Wikimedia Commons

11.長谷川 龍樹　他、実証的研究の事前登録の現状と実践　2021 年 92 巻 3 号 p. 188-196
https://www.jstage.jst.go.jp/article/jjpsy/92/3/92_92.20217/_article/-char/ja/

12. Michael Bosnjak , et al,　A Template for Preregistration of Quantitative Research in Psychology: Report of the Joint Psychological Societies Preregistration Task Force,　https://osf.io/preprints/psyarxiv/d7m5r_v1

13. Michael Bosnjak , et al,　https://syasin.hus.osaka-u.ac.jp/Bosnjak_et_al(2021)JPtranslation.pdf

14. Nature Human Behaviour. (2020, July 30).　ありのままを伝えよ。 (樋口匡貴, 三浦麻子, & 藤島喜嗣, Trans.). note. https://note.com/socpsy/n/nfdeb4538fb12

15. PCI-RRのポリシー　2.1 Managing Board and Governance,
https://rr.peercommunityin.org/about/full_policies#h_43228551620616133073406 46

16. 水野景子, & 清水裕士. (2023). 社会的ジレンマ状況における罰の逆効果の検討：Remove the sanction パラダイムを用いて. 社会心理学研究, 38(3), 51–58. https://doi.org/10.14966/jssp.2205

17. Yamada, Y. (2018). How to Crack Pre-registration: Toward Transparent and Open Science. *Frontiers in Psychology*, *9*, 1831. https://doi.org/10.3389/fpsyg.2018.01831

18. Christopher D. Chambers & Loukia Tzavella, The past, present and future of Registered Reports. Nature Human Behaviour volume 6, pages29–42 (2022). https://www.nature.com/articles/s41562-021-01193-7

19. Science must respect the dignity and rights of all humans. (2022). Nature Human Behaviour, 6(8), 1029–1031. https://doi.org/10.1038/s41562-022-01443-2

20. Williams, L. E., & Bargh, J. A. (2008). Keeping one's distance: the influence of spatial distance cues on affect and evaluation. Psychological Science, 19(3), 302–308. https://doi.org/10.1111/j.1467-9280.2008.02084.x

21.藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における"p-hacking"の実践例. 心理学評論, 59(1), 84–97. https://doi.org/10.24602/sjpr.59.1_84

22.藤島喜嗣, & 樋口匡貴. (2016). 社会心理学における"p-hacking"の実践例. 心理学評論, 59(1), 84–97. https://doi.org/10.24602/sjpr.59.1_84