Replikatsioonikriisist

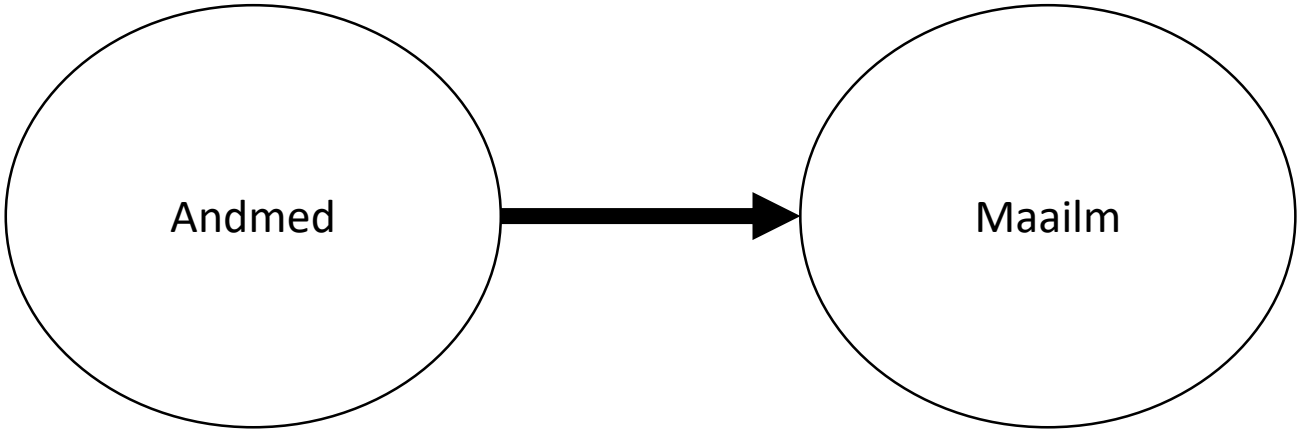
ja selle mõjust teadusele väikses kogukonnas

Peeter Tinits

Teoreetiline keeleteadus Eestis V

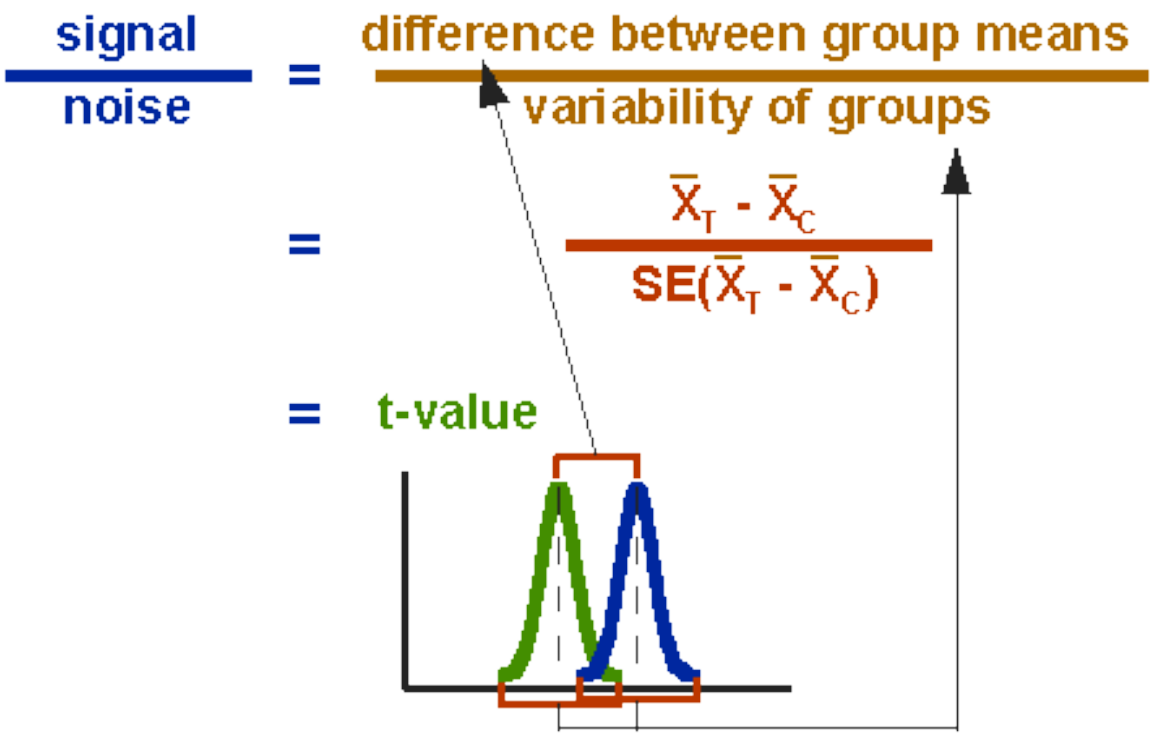
24.11.2017

Mida me uurime?



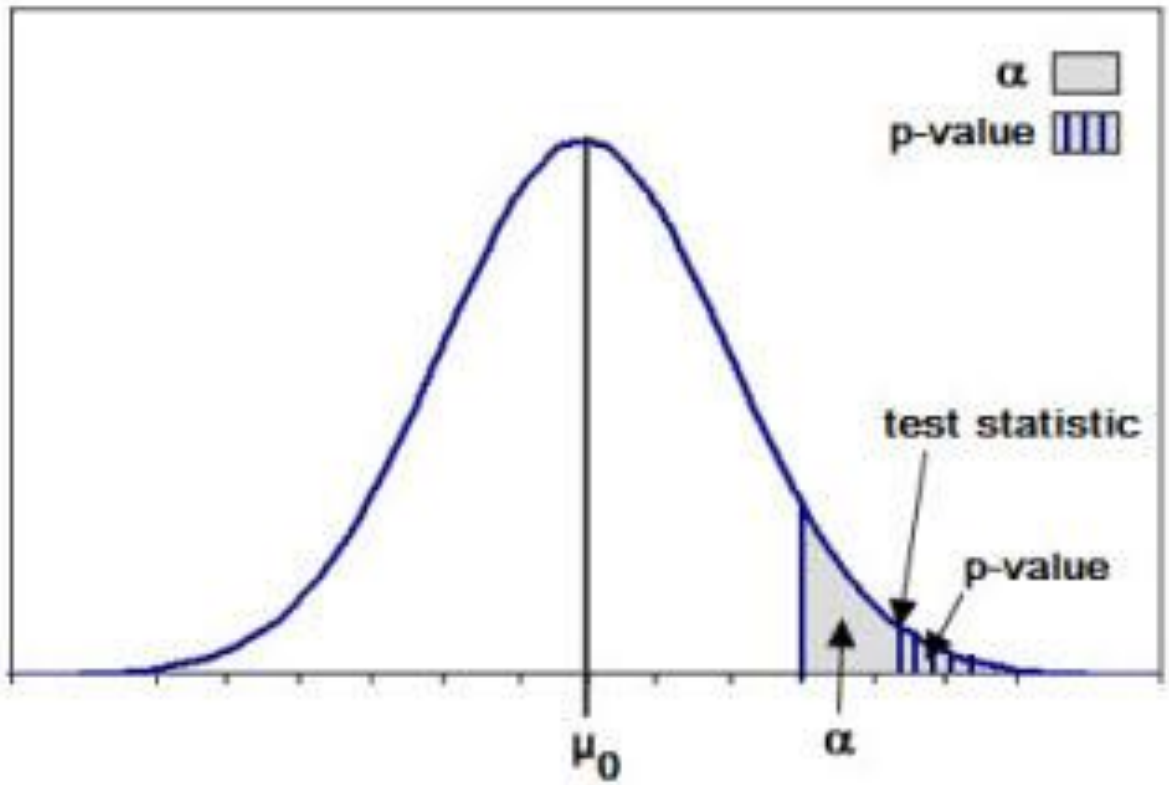
Hüpoteesi testimine

- Nt t-test
  - 1) Arvutame teststatistiku



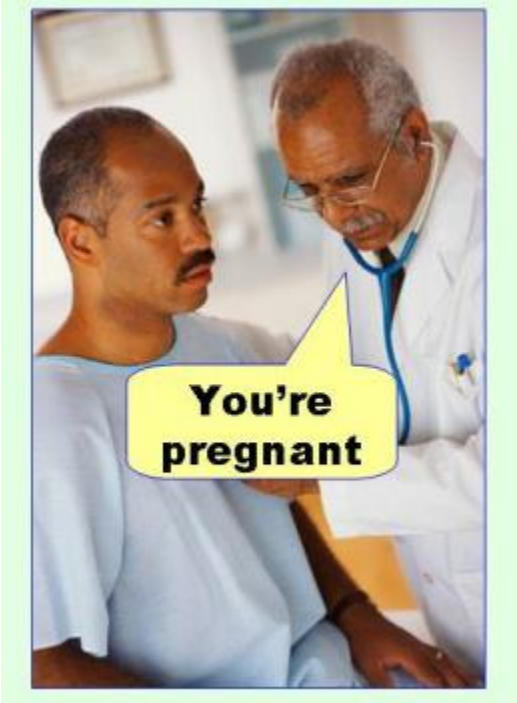
Hüpoteesi testimine

- 2) Kas test-statistik on üllatav?
  - p < .05 = jah/ei

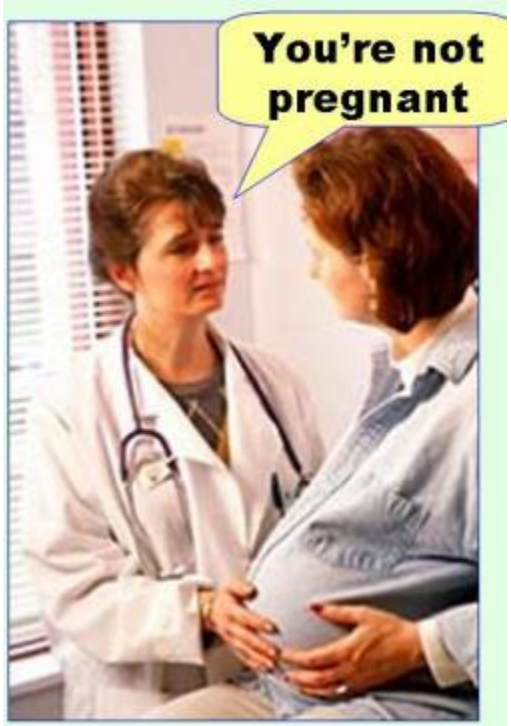


Eksimisvõimalused

- Tüüp I viga
- (vale kinnitus, *false positive*)



- Tüüp II viga
- (vale ümberlükkamine, *false negative*)



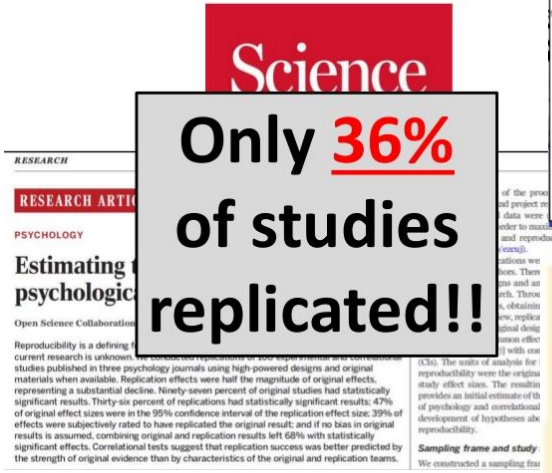
Hüpoteesid ja (vale)leiud

Ioannidis (2005) [Why Most Published Research Findings Are False.]

Mark Andrews. 2017. [False discovery app.]

Vt veel [https://lawsofthought.github.io/replication-crisis-demos/]

Replikatsioonikriis teaduses



Only 36% of studies replicated!!

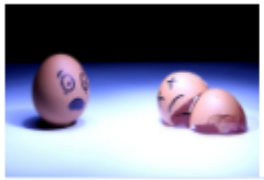
"There is increasing concern about the reliability of biomedical research, with articles suggesting that up to 85% of research funding is wasted."

Bustin, S. A. (2015). The reproducibility of biomedical research: Sleepers awake! *Biomolecular Detection and Quantification*

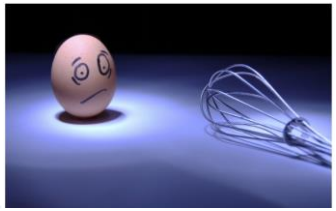
ERR novaator
KOOSTÖÖS TARTU ÜLIKOOLIGA

Katsete korratavus kütab psühholoogias kirgi

Analüüs: teaduskirjandus on kiivas, kriisist pole mõtet rääkida



Enam kui poolte psühholoogiakatsete tulemused ei pruugi vett pidada



Kuidas ära tunda usaldusväärset teadustööd?

Replikatsioonikriis teaduses

- ... (ja varemgi)
- 1967 – „Tark uurija võib teha pikka aega näivalt mõistlikke eksperimente oma teooriat tegelikult testimata“ (Meehl)
- 2005 - „Miks enamik avaldatud uurimustest on valed“ (Ioannidis)
- 2008 - „Voodoo-korrelatsioonid neuroteadustes“ (Yul et al.)
- 2011
  - Simmons et al. „Valepositiivne psühholoogia“, „uurija vabaduseastmed“ (*researcher degrees of freedom*),
  - Gelman & Loken „Hargnevate teede aed“
  - Mitmed skandaalid uskumatute tulemuste ja andmete võltsimisega (Diederik Stapel, Marc Hauser, Daryl Bem)

Via Gelman 2016

Replikatsioonikriis teaduses

- 2013 – Hulk nõrku artikleid võetakse ette
  - käte paksus ja poliitiline hoiak, shokolaadi söömine ja Nobeli preemiad, mõjukate eksperimentide korduskatsed kukuvad läbi
- 2015 – „Jõupooside“ analüüsi kriitika, suurte replikatsiooniprojektide tulemused

- 2016 – „Kuningas on alasti“
  - - kuulsate psühholoogiaeksperimentide korduskatsete läbikukkumine ei üllata enam
  - 70% 1500-st teadlasest tunnistavad, et on proovinud vähemalt korra ja pole õnnestunud (ajakirja *Nature* küsitlus)
  - 90% 1500-st arvavad, et on kriis (ajakirja *Nature* küsitlus)
- 2017
  - Manifestid läbipaistvama ja reprodutseeritava teaduse jaoks, märgid, programmid
  - „Viime replikatsiooni peavoolu“ jne.

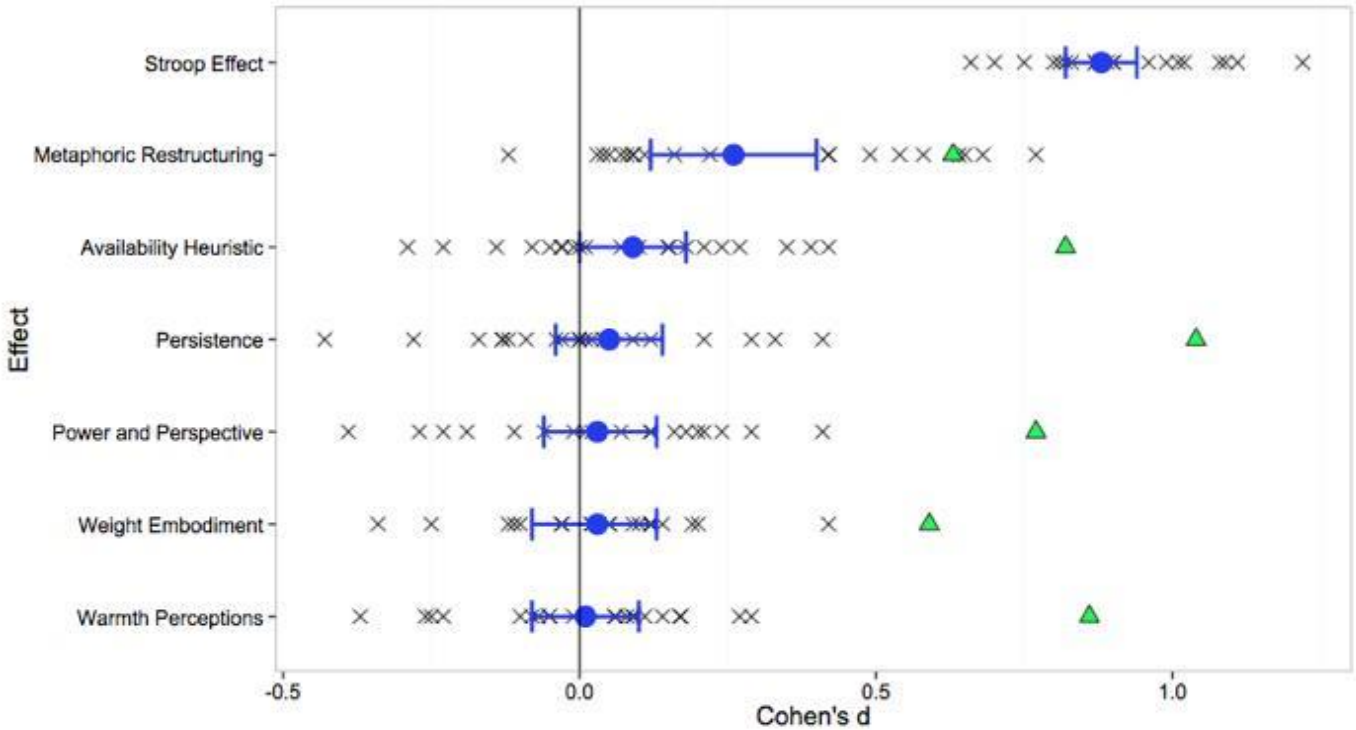
Via Gelman 2016 jt

Replikatsioonikriis

Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

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

Replikatsioonide tulemused



Many Labs 3, 2015, https://osf.io/tukby/, https://osf.io/j9ady/

Leviv arusaam

„[...] suur osa teaduskirjandusest, võib-olla pool sellest, võib osutuda lihtsalt valeks.“

Dr. Richard Horton,
teadusliku meditsiiniajakirja Lancet peatoimetaja (2015)

„[...] ei ole lihtsalt enam võimalik uskuda suurt osa kliinilistest uurimustest [...]“

Dr. Marcia Angell,
teadusliku meditsiiniajakirja New England Medical Journal peatoimetaja (2015)

P-häkkimine

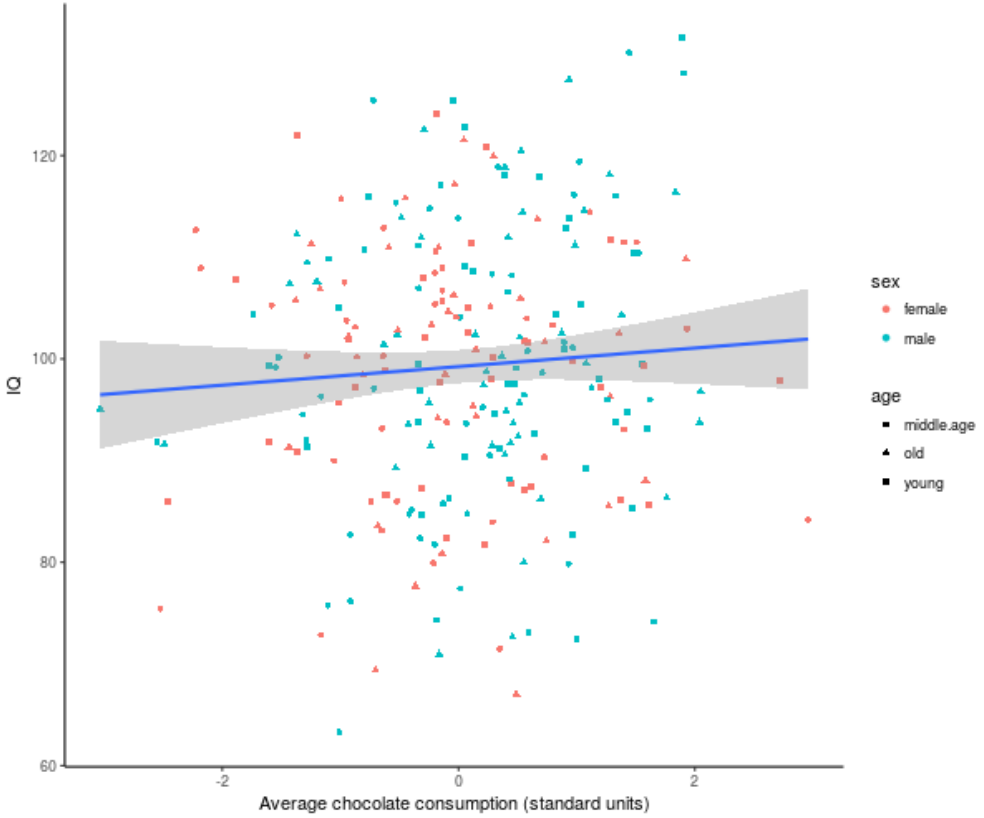
P-hacking to achieve significant results

Select which subgroups to include in an analysis, and watch the p-value change. This is one of many Questional Research Practices (QRPs) that are known as p-hacking. You can read more about p-hacking in the Simmons, Nelson & Simonsohn (2011) paper False Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. This demo is with artificial data, but a more extensive demo, using real world data, is provided by fivethirtyeight.com: Hack Your Way To Scientific Glory. All code for this demo can be found on GitHub.

Select subgroups to include:

- [x] Males
- [x] Females
- [x] Young
- [x] Middle aged
- [x] Old

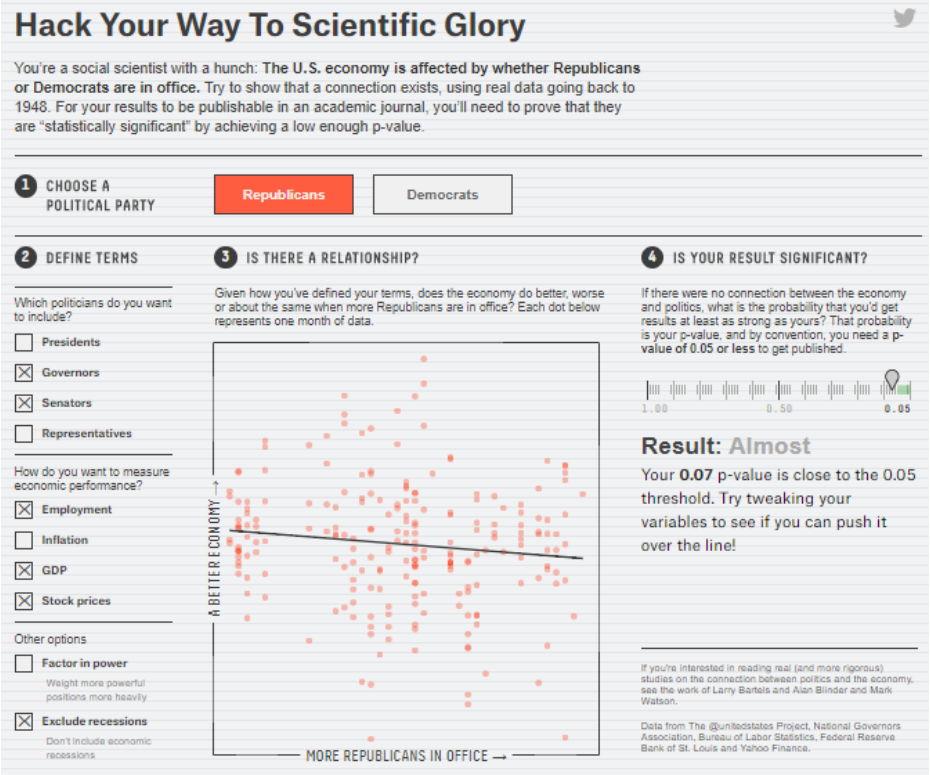
We hypothesize that chocolate lovers have higher IQs. So we collect some data.



Mark Andrews. 2017. P-hacking app.

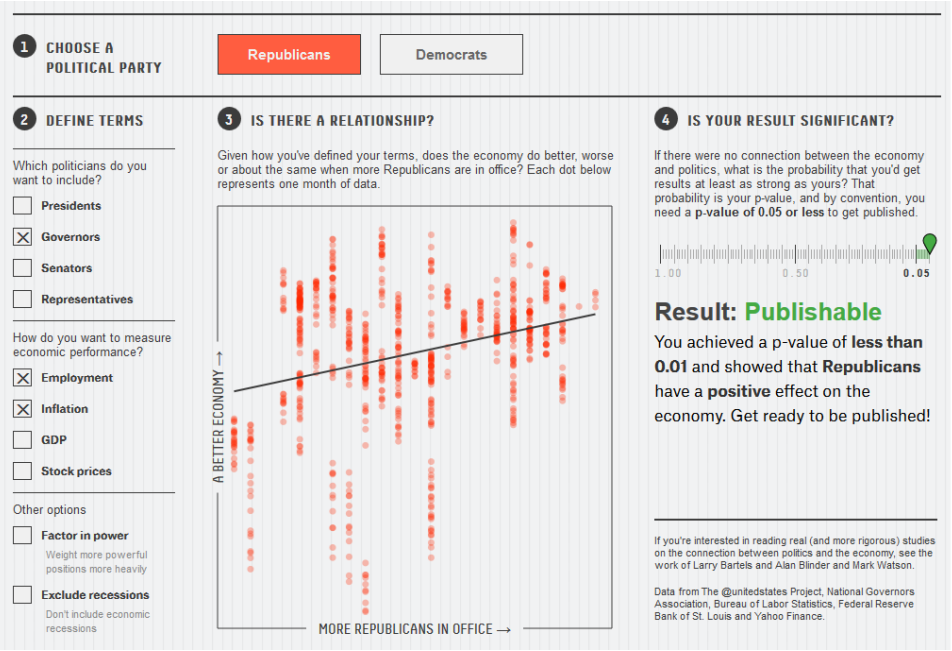
Vt veel https://lawsofthought.github.io/replication-crisis-demos/

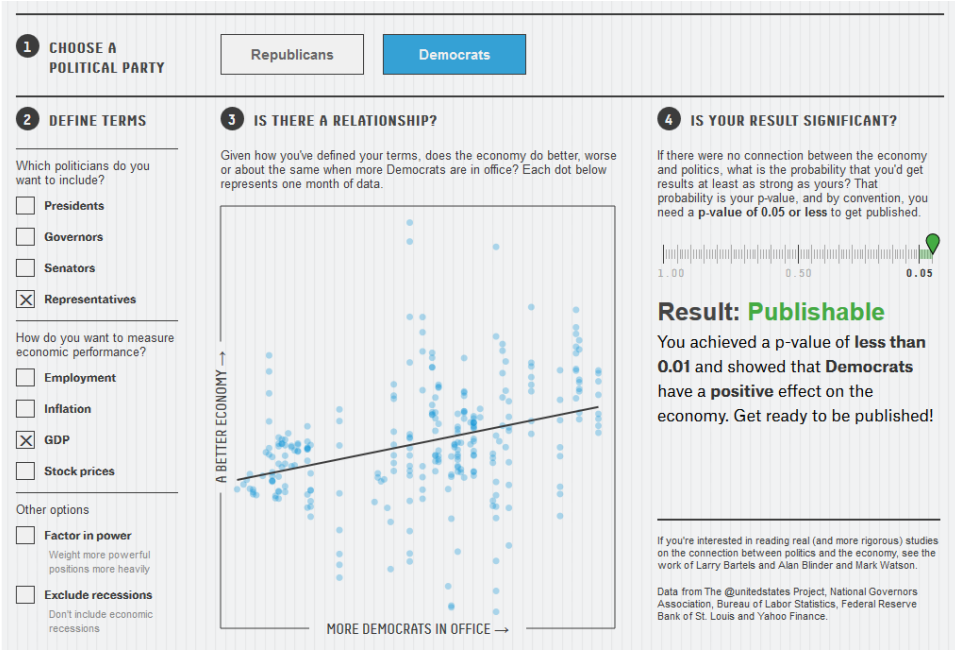
P-häkkimine



Aschwanden 2015. Science isn't broken: It's hell of a lot harder than we give credit for
https://fivethirtyeight.com/features/science-isnt-broken/
https://projects.fivethirtyeight.com/p-hacking/

P-häkkimine

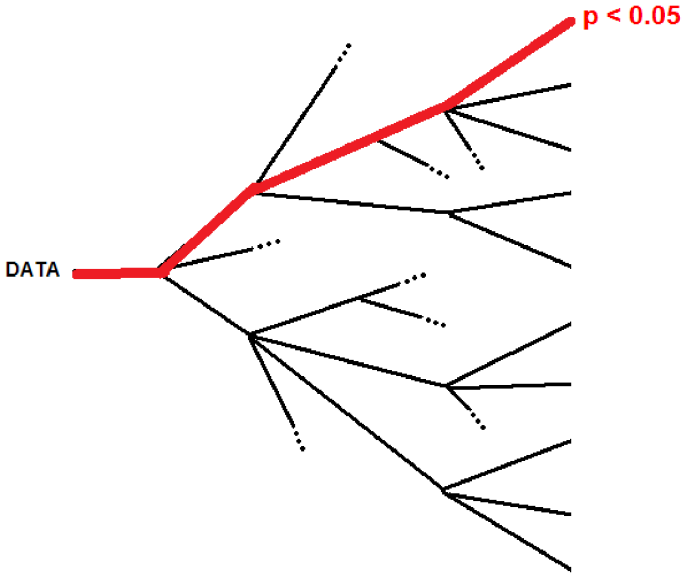




Aschwanden 2015. Science isn't broken: It's hell of a lot harder than we give credit for
https://fivethirtyeight.com/features/science-isnt-broken/
https://projects.fivethirtyeight.com/p-hacking/

Hargnevate teede aed

- Jorge Luis Borgese novelli järgi

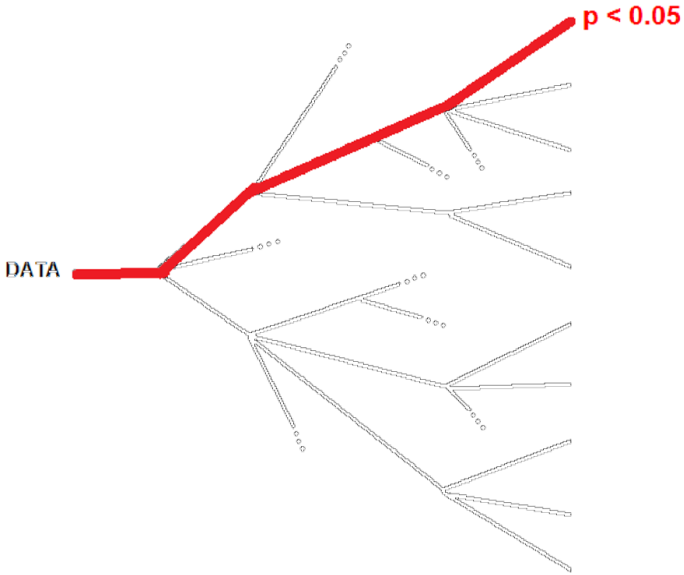


Pilt Neuro Skeptic (2015)

Andrew Gelman & Eric Loken 2013. [The garden of forking paths: Why multiple comparisons can be a problem, even when there is no ``fishing expedition'' or ``p-hacking'' and the research hypothesis was posited ahead of time]. (Unpublished.

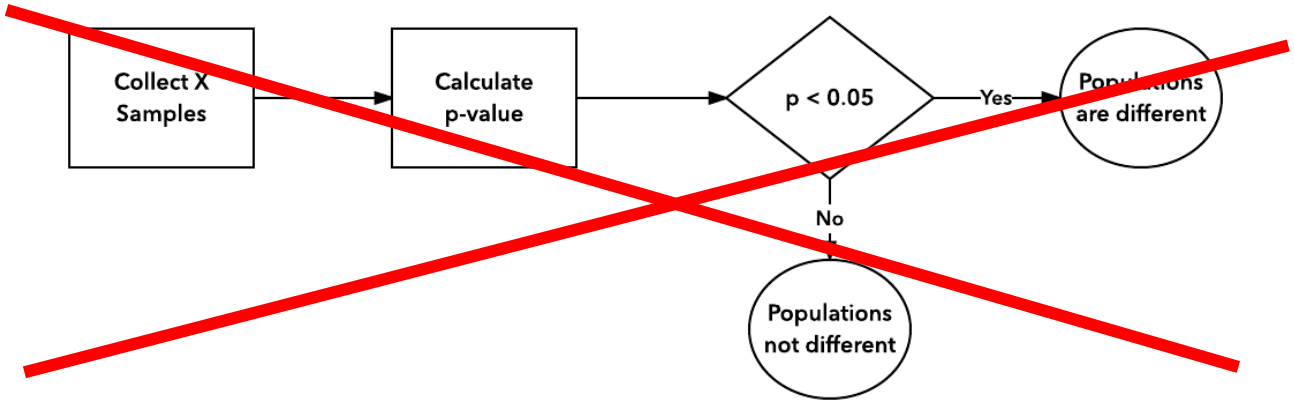
Oletame, et läksime otse

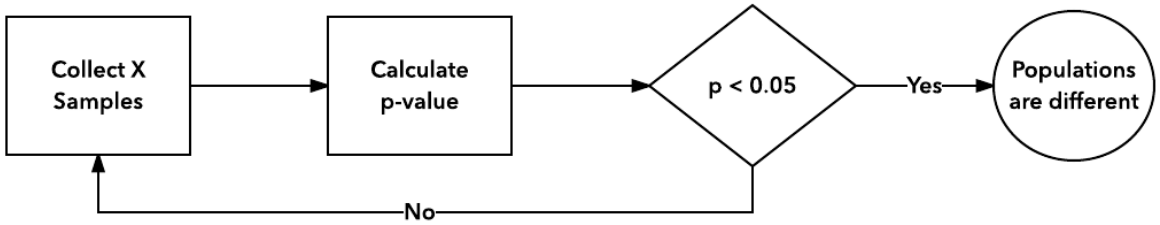
- <u>Kirjeldus:</u>
- Me arvasime, et tee on p <.05
- Leidsime p <.05
- !!!



Palju erinevaid viise

- Viis 2: osalejate lisamine

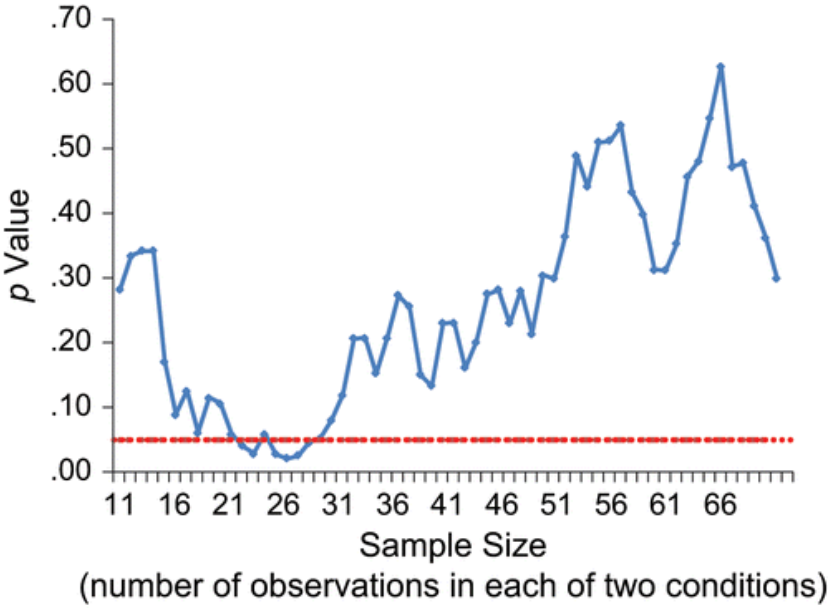




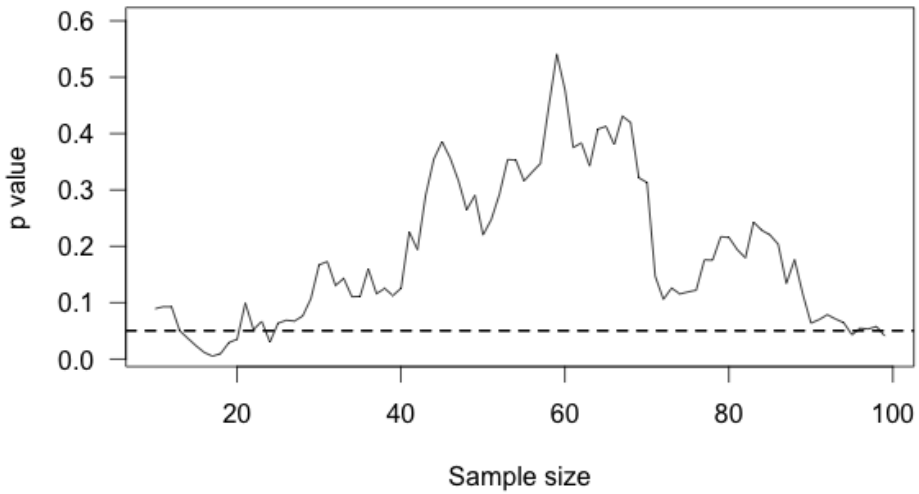
Pilt Draper 2016

Osalejate lisamine

- Lisame osalejaid ja testime iga kord
- Juhus viib p-väärtuse üle lävendi



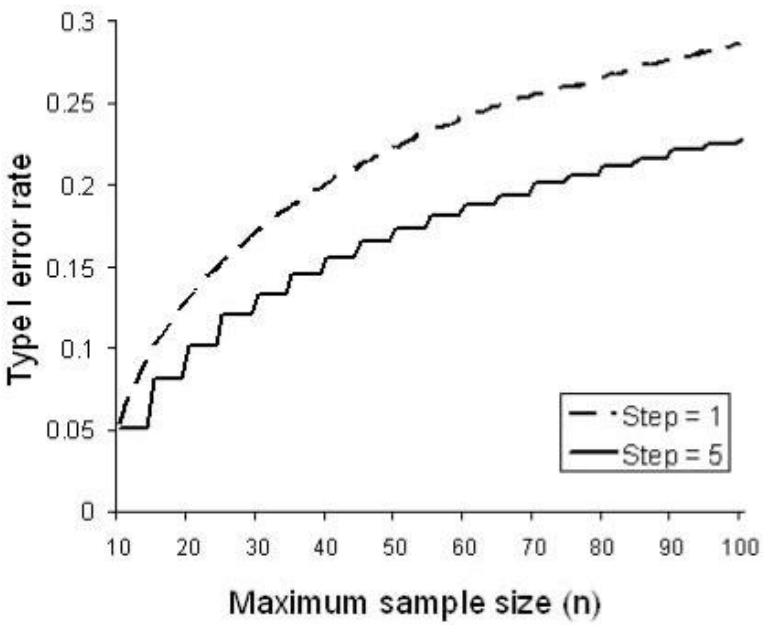
Simmons et al. 2011 [False positive psychology]



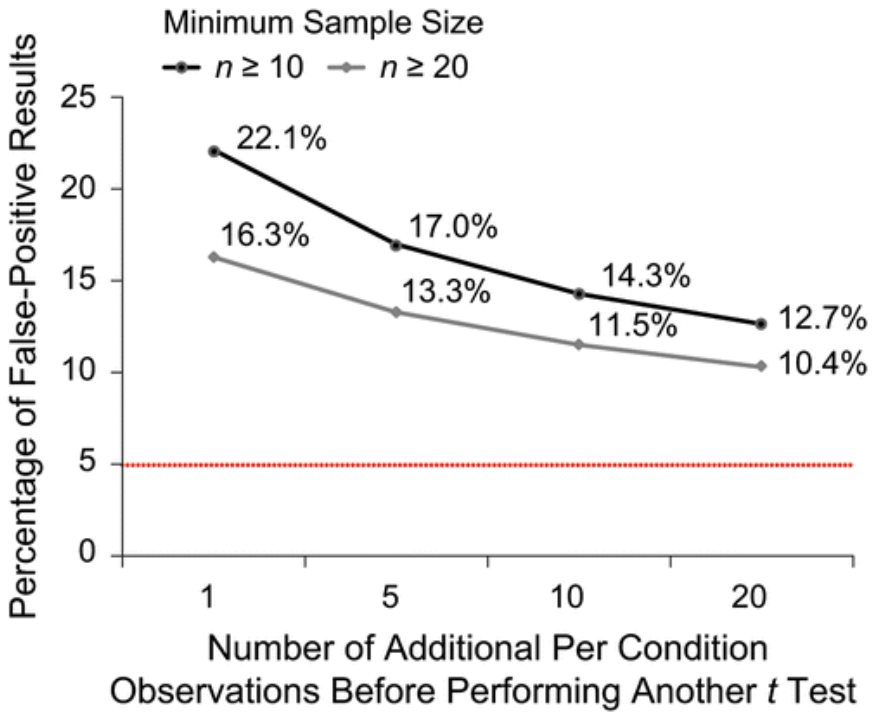
Reinhart 2015. [Statistics done wrong]

Andmestiku suurendamine

- Lisame ja testime uuesti, jne.



Yarkoni & Braver 2010
why data peeking is evil

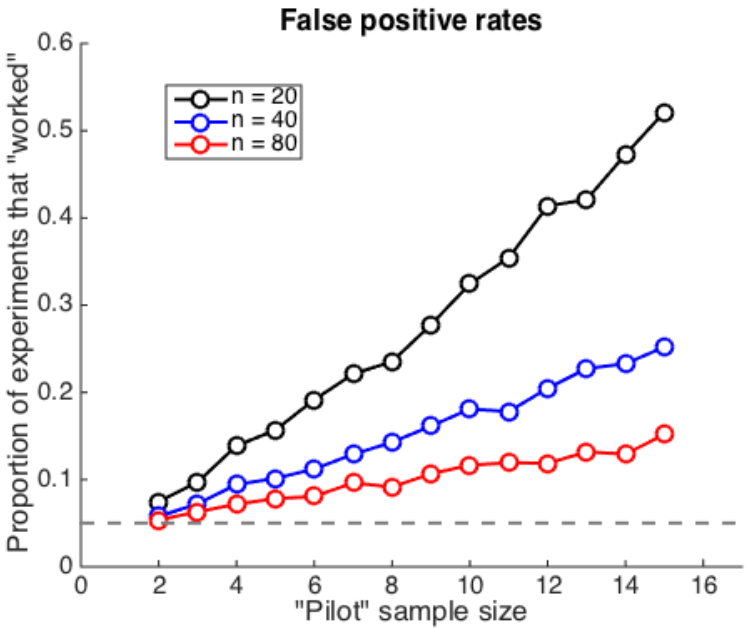


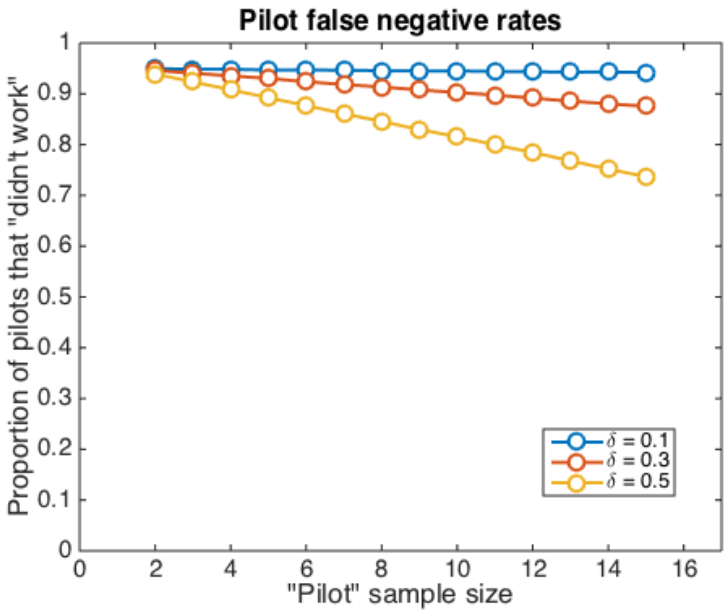
Simmons et al. 2011 False positive psychology

Vt ka https://lawsofthought.shinyapps.io/optional_stopping/

Pilootuuringud

- Kannatab ka katse võimekus leida tõest mõju





Schwarzkopf 2016. On the worthlessness of inappropriate piloting
Schwarzkopf 2016. On the magic of independent piloting

Kuidas saada p<.05

- Lõpeta andmete korjamine kui p<.05
- Analüüsi mitmeid mõõdikuid, aga kirjelda ainult neid, kus p<.05
- Kogu ja analüüsi mitmeid katsetingimusi, aga kirjelda ainult neid, kus p<.05
- Kaasa lisaparameetreid, et saada p<.05
- Jätta katsealuseid kõrvale, et saada p<.05
- Muunda andmete kuju ja skaalat, et saada p<.05

Nelson 2014. False-Positives, p-Hacking, Statistical Power, and Evidential Value

Näidis

- Kuidas üht laulu (vs teist laulu) kuulates võid
  - Tunduda endale vanem
  - Ollagi päriselt vanem

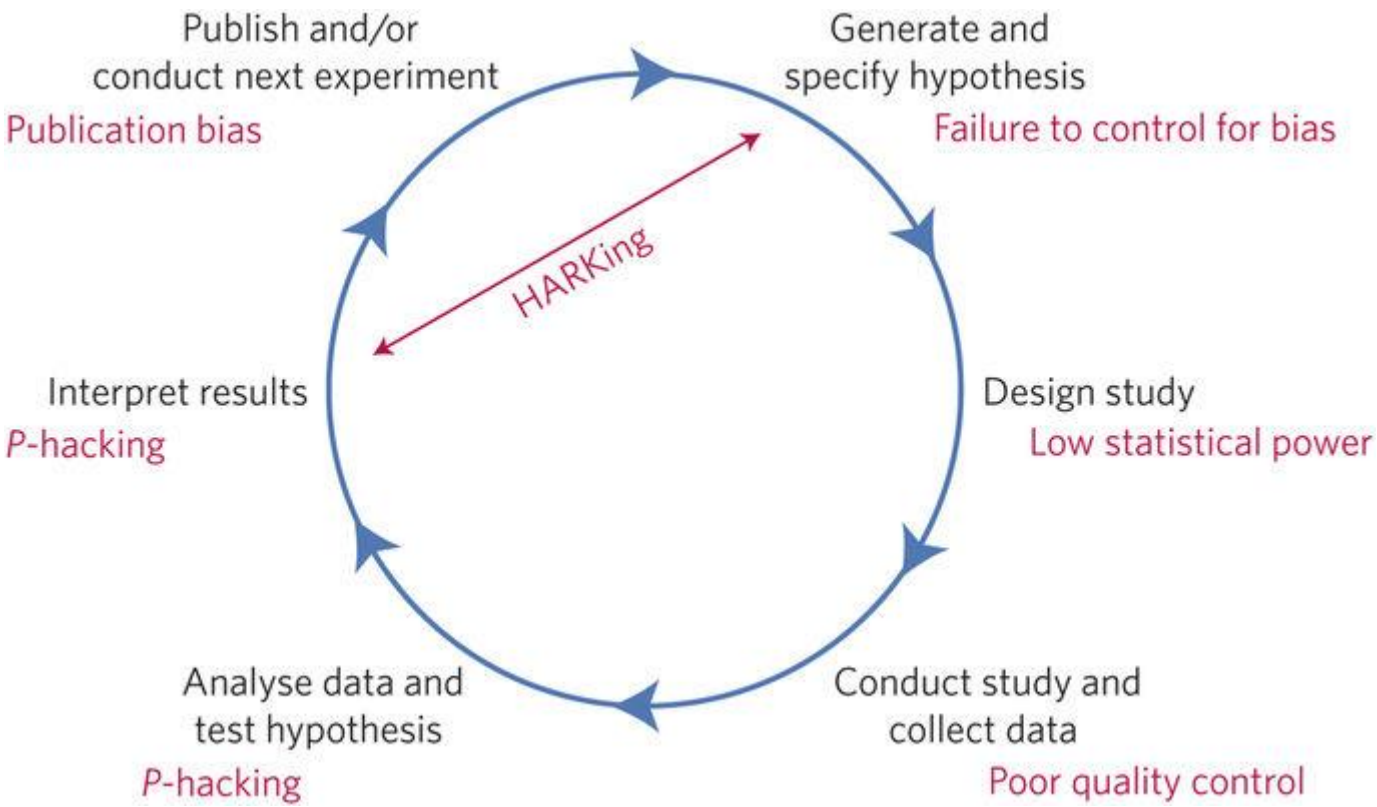
Table 3. Study 2: Original Report (in Bolded Text) and the Requirement-Compliant Report (With Addition of Gray Text)

Using the same method as in Study 1, we asked 20 34 **University of Pennsylvania undergraduates to listen** only **to either "When I'm Sixty-Four" by The Beatles or "Kalimba"** or "Hot Potato" by the Wiggles. We conducted our analyses after every session of approximately 10 participants; we did not decide in advance when to terminate data collection. **Then, in an ostensibly unrelated task, they indicated** only **their birth date (mm/dd/yyyy) and** how old they felt, how much they would enjoy eating at a diner, the square root of 100, their agreement with "computers are complicated machines," **their father's age**, their mother's age, whether they would take advantage of an early-bird special, their political orientation, which of four Canadian quarterbacks they believed won an award, how often they refer to the past as "the good old days," and their gender. **We used father's age to control for variation in baseline age across participants.**

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted M = 20.1 years) rather than to "Kalimba" (adjusted M = 21.5 years), $F(1, 17) = 4.92$, $p = .040$. Without controlling for father's age, the age difference was smaller and did not reach significance (Ms = 20.3 and 21.2, respectively), $F(1, 18) = 1.01$, $p = .33$.

Simmons et al. 2011 [False positive psychology]

Uurimistsükkel



Munafo et al. 2017. [A manifesto for reproducible science]

Ongi kõik?

Image from Naro 2016. [Repeat after me]

Lahendused

- Rohkem reegleid
  - Eelregistreerimine, uurija vabaduse astmete jäädvustamine, parem teadvustatus
- Rohkem avatust
  - Andmestikud, töökäikude jagamine, tulemuste vastastikune kontrollimine
- **Rohkem teooriat!**
  - Arvud üksi ei maksa palju, ka üks uurimus ei maksa palju

Spekulatsioone

- Spekulatsioone (Lakens 2017):
  - Replikatsioonikriis [2011-2017]
  - **Teooriakriis [2017-2025]**
  - Falsifitseerimise kriis [2025-2030]
  - Mõõtmiskriis [2030-2036]
  - Koostöökriis [2036-2048]
  - Psühholoogia kuldajastu [2050-praegu]
  - Selgeltnägemise avastamine

- Replication Crisis [2011-2017]
- Theory Crisis [2017-2025]
- Falsification Crisis [2025-2030]
- Measurement Crisis [2030-2036]
- Collaboration Crisis [2036-2048]
- Golden Age of Psychology [2050-now]
- Discovery of Pre-Cognition

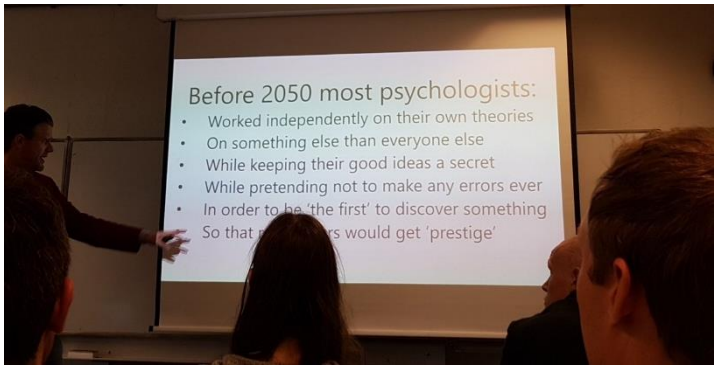
Lakens 2017. Towards a more collaborative science with StudySwap

Teooriakriis [2017-2025]?

- Teooriad on tihti ähmased
  - ei paku piisavalt mõõdetavaid küsimusi,
  - ei ennusta protsessi kulgu,
  - ei eristu teineteisest.
  - Jne

Veel spekulatsioone

- Spekulatiivne ajalugu (Lakens 2017):
  - „Enne aastat 2050, enamik psühholooge
  - töötasid iseseisvalt oma teooriate kallal,
  - igaüks omal teemal,
  - hoides häid ideid salajas,
  - teeseldes, et ei tee kunagi vigu,
  - et olla „esimesed“ kes midagi avastavad,
  - et nad seeläbi saaksid „prestiiži“.“



Daniel Lakens 2017

Mis siis selle väikse kogukonnaga?

- Samad probleemid, vähem ressursse
- Peab olema veel tublim ja täpsem
- Kontrollida ja korrata, jagada andmeid ja muresid
- => aastaks 2050 teeme kõik koostööd?

Tänan kuulamast