Präregistrierung

Weshalb und wie?

Samuel Merk - Universität Tübingen

Zeitlicher Ablauf

Intro

Inhalte des Workshops:

- Notwendigkeit der Präregistrierung:
  - Axiomatik des NHST
  - Resultierende Probleme
- Durchführung der Präregistrierung:
  - OSF: Möglichkeiten & Stolpersteine
  - OSF: Do it yourself

Umsetzung des Workshops:

- Gerne viel Eigenaktivität!
  - Produktive Heterogenität 😊
  - Hoffentlich nicht zu überformend

Kurze Survey

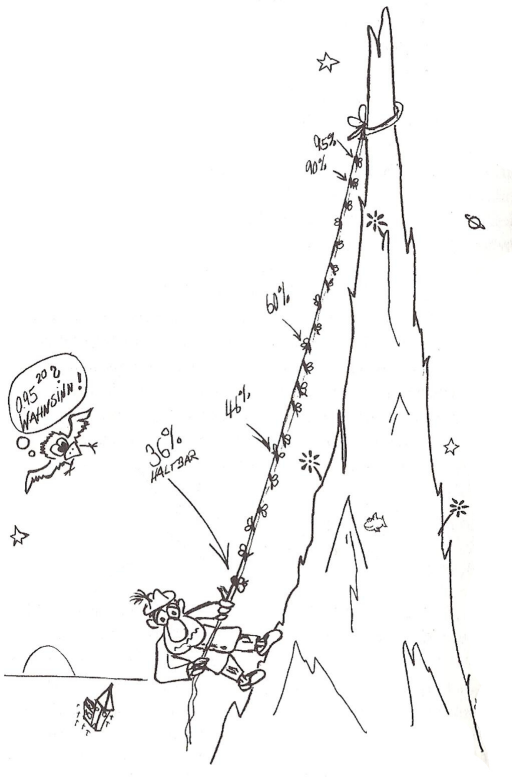
Bitte beantworten Sie die Fragen auf pollev.com/merk

Weshalb präregistrieren?

Das NHSTF

- NHSTF = Nullhypothesen Signifikanztest Framework
- p-Wert $\equiv P[E|H_0]$ wobei E als vorliegendes oder noch stärker gegen H_0 sprechendes empirisches Ereignis definiert wird *(Eid, Gollwitzer, & Schmitt, 2013)*
- Probleme:
  - Bewusstes Fehlverhalten wie "cherry-picking, data dredging, significance chasing, significance questing, selective inference, p-hacking" … (hier außer Acht)
  - Inhaltsarmut: *"parody of falsificationism in which straw-man null hypothesis A is rejected"* *(Gelman, 2016)*
  - Multiple comparisons
  - Degrees of freedom *(Simmons, Nelson, & Simonsohn, 2011)*: *"This 'what would have been done under other possible datasets' is central to the definition of p-value." The concern is […] multiple potential comparisons."* *(Gelman, 2016)*

Multiple Comparisons



https://wikis.fu-berlin.de/display/fustat

Freiheitsgrade I

Beispiele:

- Faktor- oder Summenscore?
- latent oder manifest?
- Cronbach oder McDonald?,
- mit neuem Item oder ohne?
- Operationalisierung A oder B?
- Imputationen oder FIML?
- Multivariat oder mehrfach univarat?
- 1-PL, 2-PL oder 3-PL?

Freiheitsgrade II

Beispiel aus Simmons et al *(2011)*

Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Note: The table reports the percentage of 15,000 simulated samples in which at least one of a set of analyses was significant. Observations were drawn independently from a normal distribution. Baseline is a two-condition design with 20 observations per cell. Results for Situation A were obtained by conducting three *t* tests, one on each of two dependent variables and a third on the average of these two variables. Results for Situation B were obtained by conducting one *t* test after collecting 20 observations per cell and another after collecting an additional 10 observations per cell. Results for Situation C were obtained by conducting a *t* test, an analysis of covariance with a gender main effect, and an analysis of covariance with a gender interaction (each observation was assigned a 50% probability of being female). We report a significant effect if the effect of condition was significant in any of these analyses or if the Gender × Condition interaction was significant. Results for Situation D were obtained by conducting *t* tests for each of the three possible pairings of conditions and an ordinary least squares regression for the linear trend of all three conditions (coding: low = −1, medium = 0, high = 1).

Lösungsvorschlag I: *(Simmons et al., 2011)*

Requirements for authors

1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
3. Authors must list all variables collected in a study.
4. Authors must report all experimental conditions, including failed manipulations.
5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

Lösungsvorschlag II: *(Simmons, Nelson, & Simonsohn, 2012)*

21 word solution:

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

(auch Bestandteil des Commitment to Research Transparency)

Lösungsvorschlag III: Präregistrierung

- 8-Fragenversion von aspredicted.org
- Preregistration Challenge des OSF
- Vorschlag von van't Veer und Giner-Sorolla *(A. E. van ’t Veer & Giner-Sorolla, 2016)*
- Vorschlag von Brandt et al *(Brandt et al., 2014)*

Arbeitsgruppen I (Tiere): Präregistrierungsformulare

Vorschläge für den Austausch:

- Welchen Forschungsarbeiten gehen Sie gerade nach?
- Welche Erfahrungen haben Sie bisher mit Präregistrierung gemacht?
- Welche Unterschiede zeigen die Präregistrierungsformulare?
- Welche praktischen oder theoretischen Fragen zur Präregistrierung haben Sie?
- Wie würde eine Präregistrierung für ihr aktuelles Vorhaben aussehen? Welche Schwierigkeiten würde diese Präregistrierung aufwerfen?

Das OSF

- *"The OSF is a free, open source service of the Center for Open Science. We're aligning scientific practices with scientific values by improving openness, integrity and reproducibility of research"* (https://osf.io)
- Open Science Cloud Framework mit Fokus auf
  - Workflow Management
  - File Sharing
  - Archivierung
  - (Prä-)Registrierung
- Umfassende Third-Party-Integration
  - , , , , …

Spezifische Open Science Features

- Lizenzierungen
- DOIs für Registrierungen
- PrePrint-Server
- Versionsmanagement
- Usage Analytics
- Wiki
- Sophistizierter API der programmatischen Zugriff erlaubt `{osfr}`, `{pyosf}`
- Anonymisierte View-Only-Links

Frage- und Hands-On-runde OSF

- Haben Sie Fragen zum OSF?
- Wollen Sie gleich das OSF ausprobieren?
  - Registrieren
  - Demo-Projekte anlegen
  - Demo-Projekte testen
  - Third-Parties ausprobieren
  - Registration anlegen
  - View Only Links testen
  - `{osfr}` ausprobieren

Arbeitsgruppen II (Pflanzen): Where to go next?

Diskussionsvorschläge:

- Wo sehen Sie in Ihrem konkreten Arbeitsumfeld Chancen und Probleme der Präregistrierung?
- Wo sehen Sie in Ihrem konkreten Arbeitsumfeld Katalysatoren und Stolpersteine für die Umsetzung von Präregistrierungen?

Literatur

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., … van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*(1), 217–224.

Eid, M., Gollwitzer, M., & Schmitt, M. (2013). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien* (3rd ed.). Beltz.

Gelman, A. (2016). The problems with p-values are not just with p-values: My comments on the recent ASA statement. Retrieved from http://andrewgelman.com/2016/03/07/29212/

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.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal*.

van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.