

Wissenschaftliches Arbeiten und Forschungsmethoden

Einheit 4: Sampling- und Analyseplan

02.11.2023 | Dr. Caroline Zygar-Hoffmann

Heutige Themen

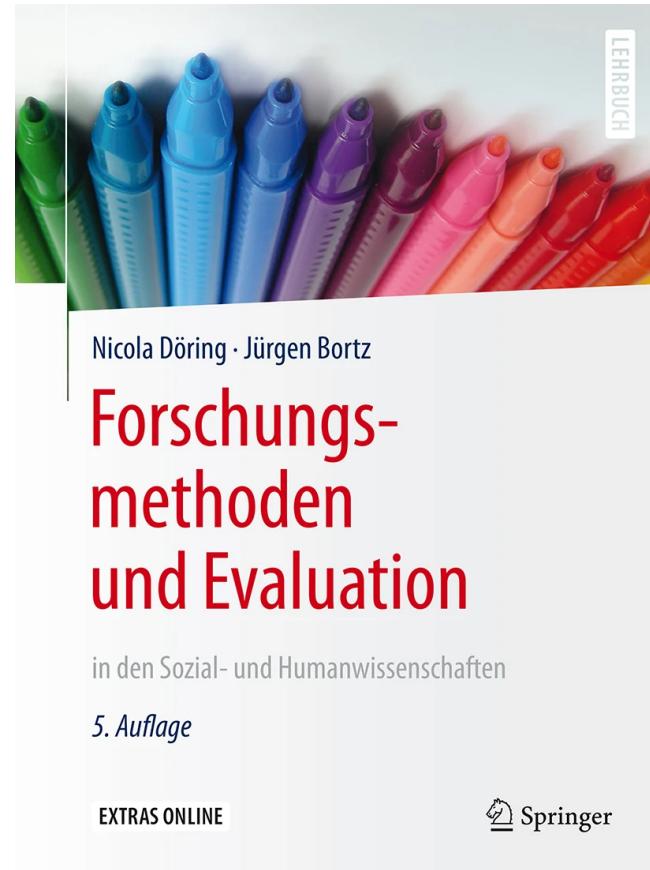
Planung von Studien und Studiendesign

Statistischer Analyseplan und Variablen

Sampling Plan (Rekrutierungsplanung)

Praxis

Literaturempfehlung für die heutige Sitzung



Kapitel 14 und 15 in Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Pearson.

Planung von Studien

Was gehört zur Studienplanung?

1. Theoriearbeit und Literaturrecherche → letzte Sitzungen
 2. Hypothesenformulierung → letzte Sitzung
 3. Design Plan → letzte Sitzung
 4. Statistischer Analyseplan → heutige Sitzung
 5. Sampling Plan (Rekrutierungsplanung) → heutige Sitzung
 6. Variablen-Operationalisierung (Auswahl Messinstrumente) → nächste Sitzung
- Studienplanung in **Präregistrierung** festhalten

Präregistrierung

M6 Masking of participants and researchers
Indicate all forms of masking and/or allocation concealment (e.g., administrators, data collectors, raters, confederates are unaware of the condition to which participants were assigned).
<u>nur falls zutreffend</u>

M11 Randomization of participants and/or experimental materials
If applicable, describe how participants are assigned to conditions or treatments, how stimuli are assigned to conditions, and how presentation of tests, trials, etc. is randomized. Indicate the randomization technique and whether constraints were applied (pseudo-randomization). Indicate any type of balancing across participants (e.g., assignments of responses to hands, etc.).
<u>nur falls zutreffend</u>

Präregistrierung

M10 Type of study and study design
Indicate the type of study (e.g., experimental, observational, <u>crosssectional</u> vs. longitudinal, single case, clinical trial) and planned study design (e.g., between vs. within subjects, factorial, repeated measures, etc.), number of factors and factor levels, etc.. in Stichpunkten

M14 Study Procedures
Please describe here any relevant information about how the study will be conducted, e.g., the number and timing of measurement time points for longitudinal research, the number of blocks or runs per session of an experiment, laboratory setting, the group size in group testing, the number of training sessions in interventional studies, questionnaire administration for online assessments, etc. in Stichpunkten

Studiendesign

Bewertungsschema

Prozedur	0	keine/unzureichende Beschreibung der Prozedur der Untersuchung, wesentliche Informationen fehlen
	1	Prozedur ist in groben Zügen beschrieben, es fehlen jedoch Informationen, so dass es nicht ohne weiteres replizierbar wäre.
	2	Prozedur ist vollständig und übersichtlich beschrieben, konzise ("auf den Punkt gebracht"). Kriterium: Sind alle Informationen vorhanden (ggf. in einem Anhang), um das Experiment replizieren zu können?

Statistischer Analyseplan

Leitfragen

- In welcher Weise sollen die elektronisch erfassten Daten vorbereitet und bereinigt werden?
- Welche statistischen Verfahren (z.B. t-Test, Regressionsanalyse, Varianzanalyse) sollen eingesetzt werden?
- Welche Statistiksoftware ist für die geplanten Analysen geeignet und zugänglich (R ist für viele Analysen ausreichend, aber für tlws. komplexe Analysen (siehe Master) weitere Software nötig, z.B. MPlus)
- Optional: Mit welchen Programmbefehlen/-paketen können die gewünschten Analysen ausgeführt werden, welche Zusatzoptionen sind wichtig (nur grob darstellen)?

Formale Voraussetzungen:

- Testauswahl nach Operationalisierung, Skalenniveau, Anzahl der Variablen und Faktorstufen
- Festlegung geplanter Kontrollvariablen
- Festlegung des Signifikanzniveaus (α -Fehler-Niveaus)
- Umgang mit fehlenden Werten und ggf. Ausreißern
- Umgang mit multiplem Testen (alpha-Fehler Kumulierung)
- ggf. präzise Angaben über praktisch bedeutsame Mindestgrößen (falls bekannt)
- Festlegung von zu berechnenden Effektstärken

Variablen

Arten von Variablen

Variablen können anhand unterschiedlicher Eigenschaften unterteilt werden:

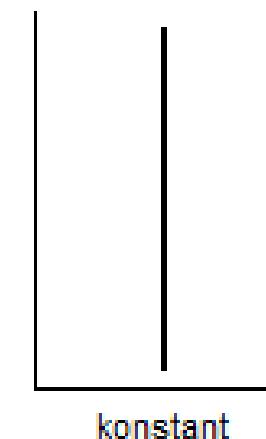
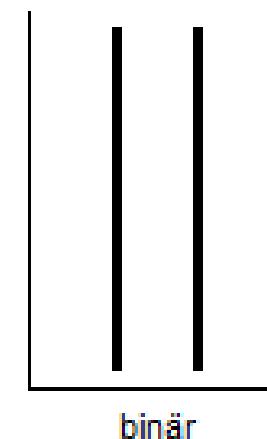
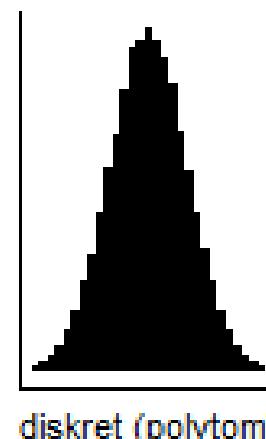
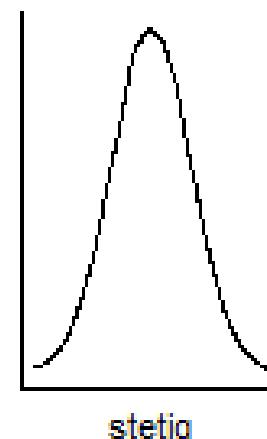
- nach Art der Merkmalsausprägungen
- nach empirischer Zugänglichkeit
- nach Stellenwert in der Untersuchung

Variablen

Arten von Variablen – nach Art der Merkmalsausprägungen

- **stetig** (kontinuierlich): jedes Intervall besitzt unendlich viele Merkmalsausprägungen (z.B. Länge, Zeit, Masse)
- **diskret** (diskontinuierlich): Intervall mit endlich vielen Ausprägungen z.B. Geschlecht, Lieblingsfarbe
 - **dichotom** (binär) = 2 Abstufungen (0, 1)
 - **polytom** = mehrfach gestuft
 - **konstant** = nur 1 Merkmalsausprägung

→ Art der Variable bestimmt das statistische Verfahren (z.B. stetiges Outcome → Regression, binäres Outcome → logistische Regression)



Variablen

Arten von Variablen – **nach empirischer Zugänglichkeit**

- **manifest** = direkt beobachtbar (Bsp. Raucher sein, Alter)
- **latent** = nicht unmittelbar beobachtbar; hypothetisches Konstrukt (Bsp. Intelligenz)

Variablen

Arten von Variablen – **nach Stellenwert in der Untersuchung**

Variablen haben im empirischen Forschungskontext unterschiedliche funktionale Bedeutungen:

- abhängige Variable (Outcome)
- unabhängige Variable (Prädiktor) *Bei manipulierten Variablen spricht man häufiger von "unabhängiger Variable", bei rein gemessenen Variablen eher von "Prädiktoren"*
- Störvariable
- Kontrollvariable / Kovariate
- Moderatorvariable
- Mediatorvariable

Variablen

Abhängige & unabhängige Variable (AV & UV)

Die Veränderung einer AV soll durch den Einfluss der UV erklärt werden.

Beispiel:

Dosis des Schlafmittel (**UV**) → Schlafdauer (**AV**)



UV gehört zum „Wenn-Teil“ bzw. dem „Je-Teil“ einer Hypothese

AV gehört zum „Dann-Teil“ bzw. „Desto-Teil“



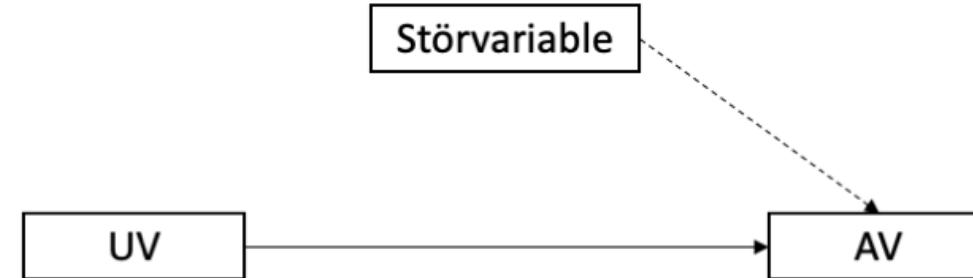
Wenn man mehr Schlafmittel nimmt, schläft man länger.



Variablen

Störvariable

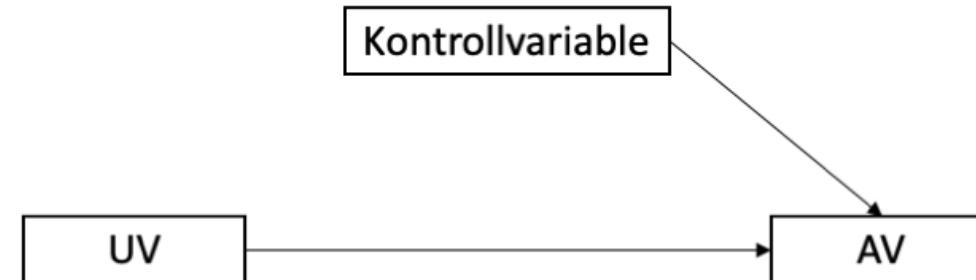
- alle Einflussgrößen auf die AV, die in einer Untersuchung nicht erfasst werden
- egal ob nicht bekannt oder vergessen



Variablen

Kontrollvariable / Kovariate

- Störvariable deren Ausprägungen erhoben (gemessen) wurde
- Einfluss kann kontrolliert wird (z.B. mittels statistischer Methoden)
- Inhaltliche Begründung (Literatur!) zur Aufnahme einer Kovariate
- Im Regressionsmodell mit inhaltlichem Prädiktor X1 und Kontrollvariable X2: Schätzung des Effekts von X1 auf Y bei Konstanthaltung des Effekts von X2 ("Kontrolle für den Einfluss von X2")



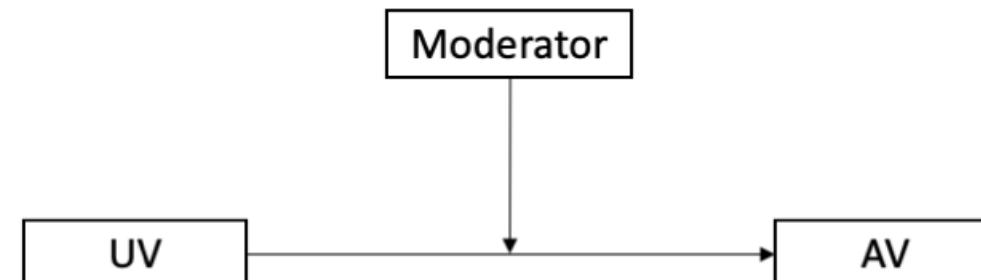
Variablen

Moderatorvariable

- **Moderator** verändert den Einfluss der UV auf die AV
- Moderationsanalyse prüft Interaktionen
- Frage: Variiert der Effekt von UV auf AV in Abhängigkeit einer weiteren Variable

Beispiel:

Schlafmitteldosis (**UV**) erhöht die Schlafdauer (**AV**); Straßenlärm (**Moderator**) verringert den Effekt der Dosis auf die **AV**



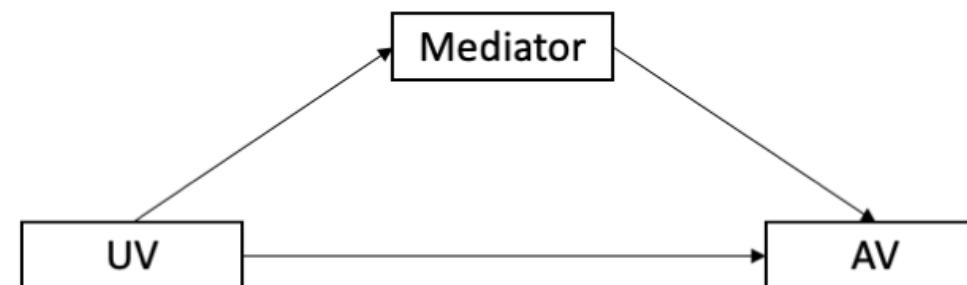
Variablen

Mediatorvariable

- **Mediator** vermittelt (**mediert**) den Einfluss der **UV** auf die **AV**
- **Indirekter Effekt:** **UV** beeinflusst **Mediator**, dies führt zur einem Effekt des **Mediators** auf **AV**
- **Direkter Effekt:** Effekt von **UV** auf **AV** (in Anwesenheit des Mediators)
- **Keine Mediation:** indirekter Effekt nicht signifikant
- **Partielle Mediation:** indirekter Effekt signifikant und direkter Effekt auch signifikant
- **Totale Mediation:** indirekter Effekt signifikant und direkter Effekt nicht mehr signifikant.

Beispiel:

Schulnote (**UV**) beeinflusst Selbstwertgefühl (**Mediator**); Selbstwertgefühl (**Mediator**) beeinflusst Lebenszufriedenheit **AV**



Variablen

Beschreibung in Präregistrierung

- Kurze Beschreibung aller gemessenen Variablen
 - abhängige Variablen (Outcomes)
 - unabhängige Variablen bzw. Prädiktoren (falls zutreffend: Moderatoren, Mediatoren)
 - Kovariate (Kontrollvariablen, die mitgemessen werden)
- Bei einem experimentellen Design alle geplanten Manipulationen sowie deren Stufen beschreiben
- Operationalisierung und Verechnungsregeln definieren (vgl. nächste Einheit)
- Abklärung der Güte von Messinstrumenten (Reliabilität, Validität, Normierung) (vgl. nächste Einheit)
- Wo zutreffend, Variablen den Hypothesen zuordnen

Präregistrierung

M12 Measured variables, manipulated variables, covariates

This section shall be used to unambiguously clarify which variables are used to operationalize the hypotheses specified above (item I3). Please (a) list all measured variables, and (b) explicitly state the functional role of each variable (i.e., independent variable, dependent variable, covariate, mediator, moderator). It is important to (c) specify for each hypothesis how it is operationalized, i.e., which variables will be used to test the respective hypothesis and how the hypothesis will be operationally defined in terms of these variables. The description here shall be consistent with the statistical analysis plans specified under AP6 (below).

in Stichpunkten

Punkt c) kann noch bis nächste Woche warten!

Präregistrierung

M7 Data cleaning and screening

Indicate all steps related to data quality control, e.g., outlier treatment, identification of missing data, checks for normality, etc.

No steps related to data quality control will be in place, so that all available data will be used (but see M8 for handling of missing data on some items of a scale).

M8 How will missing data be handled?

Indicate any procedures that will be applied during the analysis to deal with missing data, such as (a) case deletions; (b) averaging across scale items (to handle missing items for some); (c) test of missingness (MAR, MCAR, MNAR assumptions; (d) imputation procedures (FIML vs. MI); (e) Intention to treat analysis and per protocol analysis (as appropriate).

Cases will not be deleted if a participant did not finish the study, to obtain the highest sample size as possible for the analyses.

If answers on some items of a scale are missing, all answers of that scale will be treated as missing (no imputation of answers).

Präregistrierung

AP1 Criteria for post-data collection exclusion of participants, if any
Describe all criteria that will lead to the exclusion of a participant's data (e.g. performance criteria, non-responding in physiological measures, incomplete data). Be as specific as possible.
hier ggf. Ausschluss auf Basis von Manipulationscheck oder sonstiger Voraussetzung für sinnhafte Bearbeitung der Studie, ansonsten „No post-data collection exclusion of participants will be applied.“

AP2 Criteria for post-data collection exclusions on trial level (if applicable)
Describe all criteria that will lead to the exclusion of a trial or item (e.g. statistical outliers, response time criteria). Be as specific as possible.
No post-data collection exclusion on trial or item level will be applied.

Präregistrierung

AP6 Statistical models (provide for each hypothesis if varies)
<p>Specify the statistical model (e.g., t test, ANOVA, LMM) that will be used to test each of your hypotheses. Give all necessary information about model specification (e.g., variables, interactions, planned contrasts) and follow-up analyses. Include model selection criteria (e.g., fit indices), corrections for multiple testing, and tests for statistical violations, if applicable. Wherever unclear, describe how effect sizes will be calculated (e.g., for d-values, use the control SD or the pooled SD).</p>
<p><i>For H1 we will compute a directed t-test for independent samples with variable Y as outcome and variable X as group variable (group A vs. B) using the <code>t.test()</code> function in R. If variance homogeneity tested by the <code>leveneTest()</code> function from the car package is not indicated ($p < 0.05$), a Welch test will be computed instead.</i></p> <p><i>For H2 we will compute a one-way ANOVA with variable Y as outcome and variable X as group variable (group A vs. B vs. C) using the <code>lm()</code> function in R, followed by a post-hoc comparison (Tukey HSD) between the groups A and B using the <code>emmeans</code> package. Even if ANOVA assumptions are violated, we will not change the primary analysis, but discuss this as problems accordingly (and maybe explore alternative calculations).</i></p>

Präregistrierung

AP7 Inference criteria
<p>Specify the criteria used for inferences (e.g., p values, Bayes factors, effect size measures) and the thresholds for accepting or rejecting your hypotheses. If possible, define a smallest effect size of interest. If inference criteria differ between hypotheses, specify separately for each hypothesis and respective statistical model by explicitly referring to the numbers of the hypotheses. Describe which effect size measures will be reported and how they are calculated.</p>
<p>z.B.</p> <p>For H1 we will reject the H0 if the p-value of the t-test is below an alpha level of 0.05. As effect size Cohen's d calculated with the effsize package will be reported (using a pooled standard deviation).</p> <p>For H2 we will reject the H0 if the p-value of omnibus test F-test from the ANOVA is below 0.10 and the p-value of the contrast between groups A and B is below 0.10 and the effect goes in the right direction. As effect size for the ANOVA we will report eta² from the effectsize package, and for the posthoc comparison Cohen's d with the eff_size() function from the emmeans() package.</p>

Präregistrierung

AP8 Exploratory analysis (optional)

Describe any exploratory analyses to be conducted with your data. Include here any planned analyses that are not confirmatory in the sense of being a direct test of one of the specified hypotheses.

No exploration planned at this point.

Analyseplan

Bewertungsschema

Methoden	0	Es wurden keine adäquaten Methoden verwendet (z.B. nur Mittelwerte berichtet)
	2	Adäquate Test wurden gerechnet, jedoch tw. mangelhaft interpretiert ("Kausalität"), Fehler in Berechnung und/oder Interpretation, ggf. wichtige Voraussetzungen ignoriert
	4	Adäquate bzw. vereinbarte Tests wurden fehlerfrei gerechnet
	<u>Bonus:5</u>	<u>Bonus:</u> Mehr als die vereinbarten Tests gerechnet, neue Methoden selbstständig erarbeitet und korrekt angewendet. Eine explorative Analyse wurde durchgeführt, welche zum Verständnis des Phänomens beiträgt. Muss argumentativ sinnvoll eingebunden sein (einfach nur irgendetwas anderes explorativ zu rechnen reicht nicht für den Bonuspunkt).

Sampling Plan (Rekrutierungsplanung)

Auswahl der Untersuchungsobjekte

- Frage: an welchen bzw. an wie vielen Untersuchungsobjekten werden die Variablen erhoben?
- Untersuchungsteilnehmer, deren Beobachtung oder Beschreibung interessante Hypothesen versprechen

Wichtigstes Kriterium: Repräsentativität der Stichprobe

- Repräsentativität = Stichprobe muss in ihrer Zusammensetzung der Population möglichst stark ähneln
- Stichprobe ist (merkmals)spezifisch repräsentativ, wenn ihre Zusammensetzung hinsichtlich relevanter Merkmale der Populationszusammensetzung entspricht
- Stichprobe ist global repräsentativ, wenn ihre Zusammensetzung in nahezu allen Merkmalen der Populationszusammensetzung entspricht
- große Stichprobe garantiert nicht Repräsentativität (gilt nur bei unverzerrter Auswahl)
- beste Gewähr für größtmögliche globale Repräsentativität bietet die Zufallsstichprobe

Sampling Plan (Rekrutierungsplanung)

Auswahl der Untersuchungsobjekte

Einschlusskriterien:

- globale Population aller Menschen ist nicht für jede Fragestellung geeignet
- Selektion zur Erhöhung der Auftretenswahrscheinlichkeit relevanter Merkmale (z.B. bestimmte Diagnose, bestimmte Altersklasse...)

Ausschlusskriterien:

- Manche Charakteristika machen Untersuchung einzelner Personen inhaltlich/ethisch unmöglich
- Ausschlusskriterien müssen gut überlegt und genau dokumentiert werden

→ Für unser Praxisprojekt gibt es nur ein Einschlusskriterium, welches nicht inhaltlich begründet ist:
Studierender dieser Vorlesung sein. Fehlende Repräsentativität ist dann Teil der Diskussion.

Sampling Plan (Rekrutierungsplanung)

Größe der Stichprobe

- Die Festlegung des Stichprobenumfanges sollte ebenfalls in der Planungsphase erfolgen
- Verbindliche Angaben zum nötigen Stichprobenumfang für eine bestimmte Analyse nur möglich, wenn eine hypothesenprüfende Untersuchung mit vorgegebener Effektgröße geplant wird → **a priori Poweranalyse**
- Durchführung z.B. in R oder freies Programm GPower; Link zu Erklärvideo für G*Power
- Für die Größe von Stichproben, mit denen keine spezifischen Hypothesen geprüft werden oder keine Effektgröße schätzbar ist, gibt es keine genauen Richtlinien ("more is better, but be mindful of resources")
- Falls die Stichprobengröße aus anderen Gründen festgelegt ist (wie bei uns), kann man eine "**(post hoc) Sensitivitäts-Poweranalyse**" machen, um abzuschätzen wie hoch die statistische Power für die Aufdeckung eines Effekts mit dieser Stichprobengröße ist (wichtig: eine post hoc Analyse auf dem *gefundenen* Effekt aus der eigenen Studie macht keinen Sinn, nur eine auf dem erwarteten Effekt)

Sampling Plan (Rekrutierungsplanung)

Größe der Stichprobe

- Ziel der a priori Poweranalyse: Bestimmung einer Stichprobengröße, die bei angenommender Effektgröße eine statistisch-signifikantes Ergebnis ermöglicht
- Zusammenspiel aus Signifikanzniveau (Wahrscheinlichkeit für Fehler 1. Art) und Power (1 - Wahrscheinlichkeit für Fehler 2. Art)
- Poweranalyse richtet sich nach der Hauptfragestellung (primärer zu überprüfender Hypothese); bei mehreren Primärhypothesen, die Poweranalyse für den kleinsten erwarteten Effekt

Relevante Parameter:

- Effektstärke (aus Vorstudien/Literatur oder basierend auf Plausibilitätsannahme) → je größer, desto kleineres N benötigt
- Signifikanzniveau (i.d.R. alpha = .05) → je kleiner, desto kleineres N benötigt
- Power (i.d.R. mindestens 0.8) → je größer, desto kleineres N benötigt

Sampling Plan (Rekrutierungsplanung)

Poweranalyse in R

Poweranalyse für einen unabhängigen t-Test (auch für andere Hypothesentests möglich):

```
> library(pwr)
> pwr.t.test(d = 0.5, sig.level = 0.05, power = 0.80, type = "two.sample", alternative = "greater")
Two-sample t test power calculation
n = 50.1508
d = 0.5
sig.level = 0.05
power = 0.8
alternative = greater
NOTE: n is number in *each* group
```

- Bei a priori Annahme einer Effektstärke von Cohen's $d = 0.5$ für den Mittelwertsunterschied (mittlerer Effekt)
- und einem Signifikanzniveau von $\alpha = .05$
- und einer Power von 0.8 (Chance auf positives Ergebnis)

benötigt man $N = 102$ Personen (51 pro Gruppe), um in einem einseitigen (*alternative = "greater"*) unabhängigen t-Test (*type = "two.sample"*) einen signifikanten Gruppenunterschied nachzuweisen

Sampling Plan (Rekrutierungsplanung)

Sensitivitäts-Poweranalyse in R

Sensitivitäts-Poweranalyse für einen unabhängigen t-Test (auch für andere Hypothesentests möglich):

```
> library(pwr)
> pwr.t.test(d = 0.5, sig.level = 0.05, n = 17, type = "two.sample", alternative = "greater")
Two-sample t test power calculation

n = 17
d = 0.5
sig.level = 0.05
power = 0.4137168
alternative = greater

NOTE: n is number in *each* group
```

- Bei a priori Annahme einer Effektstärke von Cohen's $d = 0.5$ für den Mittelwertsunterschied (mittlerer Effekt)
- und einem Signifikanzniveau von $\alpha = .05$
- und $N = 34$ Personen (17 Personen pro Gruppe)
(Hinweis: Wir sind insgesamt 37 Studierende + 1 Dozentin. Wenn die eigene Gruppe = 3-5 Personen nicht an der Studie teilnehmen, aber sonst alle, kommen wir auf 33-35 Versuchspersonen.)

erreicht man eine Power von 41% für einen einseitigen, unabhängigen t-Test; d.h. die Wahrscheinlichkeit einen Effekt zu finden, wenn er da ist, beträgt 41%.

Sampling Plan (Rekrutierungsplanung)

Poweranalyse in R: Befehle

- **pwr.t.test()** für Einstichproben-t-test, sowie unabhängige und abhängige Zweistichproben-t-tests mit gleicher Gruppengröße; Angabe von Effektstärke d (s.o.)
- **pwr.anova.test()** für einfaktorielle ANOVA; Angabe von Effektstärke f
 - $f = \sqrt{(\omega^2/1 - \omega^2)}$
 - Angabe von k = Stufen des Faktors
- **pwr.f2.test()** für Regression und mehrfaktorielle ANOVA; Angabe von Effektstärke f^2
 - $f^2 = R^2/(1-R^2)$ bzw. $\omega^2/1 - \omega^2$ für aufgeklärte Varianz des Gesamtmodells
 - $f^2 = (R^2_{AB} - R^2_A)/(1-R^2_{AB})$ bzw. $\omega^2_{partial_B}/1 - \omega^2_{partial_B}$ für aufgeklärte Varianz des Prädiktors B
 - Angabe von u = Anzahl der Prädiktoren/Faktoren ohne Intercept
 - Rückgabewert v + u + 1 ergibt Stichprobengröße n
- **pwr.r.test()** für Korrelation

Sampling Plan (Rekrutierungsplanung)

Poweranalyse: Woher weiß ich die Effektstärke?

- **Recherche in bestehender Literatur!** Diese dann herunterkorrigieren (wegen Publication Bias), d.h. erwarteten Effekt für die Poweranalyse kleiner angeben, als das was in der Literatur gefunden wurde.
- Falls nötig, **Transformation der Effektstärke** aus der Literatur in die Effektstärke, die für die Poweranalyse gebraucht wird (z.B. r in d):
https://www.psychometrica.de/effect_size.html
- Alternativen: **Pilotstudien** (aber auch nicht zuverlässig) oder **praktisch bedeutsame Mindestgrößen**
- Bei unklarer zu erwartender Effektstärke:
Mehrere Poweranalysen rechnen, visualisieren wie sich die nötige Stichprobenzahl abhängig von der Effektstärke verändert ("Power curves")

14. Transformation of the effect sizes d , r , f , Odds Ratio, η^2 and Common Language Effect Size (CLES)

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239), Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios) and Dunlap (1994; transformation in CLES).

Effect Size	<input type="button" value="d"/>	0.5
d	<input type="text" value="0.5"/>	0.5
r	<input type="text" value="0.2425"/>	0.2425
η^2	<input type="text" value="0.0588"/>	0.0588
f	<input type="text" value="0.25"/>	0.25
Odds Ratio	<input type="text" value="2.4766"/>	2.4766
Common Language Effect Size CLES	<input type="text" value="63.82 %"/>	63.82 %
Number Needed to Treat (NNT)	<input type="text" value="3.6189"/>	3.6189

Remark: Please consider the additional explanations concerning the transform from d to Number Needed to Treat in the section **BESD and NNT**. When using r as the initial effect size, the calculator draws on the formula specified by Dunlap (1994) for the conversion to CLES: $CLES = \frac{\arcsin(r)}{\sqrt{2}} + .5$. In all other case d is applied in accordance with McGraw and Wong (1992): $CLES = \frac{\phi(d)}{\sqrt{2}}$.

Sampling Plan (Rekrutierungsplanung)

Poweranalyse: Woher weiß ich die Effektstärke?

- Durchschnittlicher Effekt in der Psychologie: $r \sim .20$ ($d \sim .40$) mit einer Standardabweichung von $r \sim .10\text{-.}15$
- Das ist kleiner als die Einordnung, die Cohen vorgeschlagen hat
- Vergleich mit gut vorstellbaren Effekten, z.B.
 - mittlerer Placebo-Effekt: $d = 0.24$
 - Veränderung der Lebenszufriedenheit direkt nach Hochzeit: $d = 0.26$ und direkt nach einem Trauerfall: $d = -0.48$
 - mittlerer Effekt des Schulbesuchs in der ersten Klasse auf die Mathefähigkeit: $d = 1.1$

Source	Small	Medium	Large
Richard et al (2003): characterizations of small, medium and large effects based on 25,000 social psychological studies	.1	.2	.3
Hemphill (2003): Lower, middle and upper thirds of 380 meta-analyses from psychological assessment and intervention literature	<.2	.2-.3	>.3
Bosco et al (2015): Lower, middle and upper thirds of 147,328 applied psychology correlations	<.09	.09-.26	>.26
Gignac & Szodorai (2016): 1st, 2nd, and 3rd quartiles from 87 meta-analyses from individual differences journals	.11	.19	.29
Paterson et al (2016): 1st, 2nd, and 3rd quartiles of correlations from 258 meta-analyses in organizational behavior	.12	.20	.31

Cohen's Einordnung:

		small	medium	large
t-test for means	d	.20	.50	.80
t-test for corr	r	.10	.30	.50
F-test for regress	f²	.02	.15	.35
F-test for anova	f	.10	.25	.40
chi-square	w	.10	.30	.50

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2, 156-168. doi:10.1177/2515245919847202

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, 100(2), 431-449. http://doi.org/10.1037/a0038047

Hróbjartsson, A., & Gøtzsche, P. C. (2004). Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *Journal of internal medicine*, 256(2), 91-100.

Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: a meta-analysis. *Journal of Personality and Social Psychology*, 102(3), 592-615. http://doi.org/10.1037/a0025948

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2, 172-177. doi:10.1111/j.1750-8606.2008.00061.x

Sampling Plan (Rekrutierungsplanung)

Poweranalyse: Wie wähle ich mein Alpha-Niveau?

- Standard in der Psychologie: $\alpha = 0.05$
- Benjamin et al. (2018): Vorschlag, das α -Niveau weiter auf 0.005 zu senken, um Replizierbarkeit der Effekte zu erhöhen
- Lakens et al. (2018): Justify your $\alpha!$
 - Für das Praxisprojekt könnte das bedeuten mal in der Sensitivitätspower-Analyse verschiedene Werte von Alpha auszuprobieren, um zu sehen wie sich die Power dadurch erhöht
 - Meine Empfehlung: Für dieses Ausprobieren α maximal so weit erhöhen wie es nicht den β -Fehler (1-Power) übersteigt
 - Begründung: Im Rahmen dieses Praxisprojekts, in dem wir nur üben, ist es genauso "schlimm" zu behaupten es gäbe einen Effekt, wenn es keinen gibt, als zu behaupten es gäbe keinen Effekt, wenn es einen gibt.
 - In der Forschung wird der Fehler 1.Art (α -Fehler) zumindest formal konstant niedrig gehalten, während die Höhe des Fehlers 2.Art (β -Fehler) maßgeblich von der Größe der Stichprobe abhängt (welche variabel ist)

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Johnson, V. E. (2018). Redefine statistical significance. *Nature human behaviour*, 2(1), 6-10.

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... & Zwaan, R. A. (2018). Justify your alpha. *Nature human behaviour*, 2(3), 168-171.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.

Sampling Plan (Rekrutierungsplanung)

Anwerbung von Untersuchungsteilnehmern:

- Wo soll rekrutiert werden?
- Mit welchen Mitteln?
- Wird es eine Vergütung geben?
 - Bei uns: Rückmeldung über eigene Antworten; bitte mitdenken und bestenfalls umsetzen! Es wird sowieso eine Übung dazu geben (d.h. machen werden Sie es sowieso), aber ob Sie es letztendlich in Ihren Fragebogen übernehmen, oder nicht, ist Ihre Sache.
- Ist die Teilnahme Teil eines erweiterten (Forschungs-/Behandlungs-)Programms

Sampling Plan (Rekrutierungsplanung)

Determinanten der freiwilligen Untersuchungsteilnahme (Bortz und Döring, 2012)

Freiwillige Untersuchungsteilnehmer ...

- ...verfügen über eine bessere schulische Ausbildung als Verweigerer (bessere Notendurchschnitte).
- ...schätzen ihren eigenen sozialen Status höher ein als Verweigerer.
- ...benötigen mehr soziale Anerkennung als Verweigerer.
- ...sind geselliger als Verweigerer.
- ...sind weniger autoritär als Verweigerer.
- ...haben eine geringere Tendenz zu konformem Verhalten als Verweigerer.
- ...geben sich in Untersuchungen über geschlechtsspezifisches Verhalten unkonventioneller.
- ...verfügen auf Basis der meisten Untersuchungsergebnisse über eine höhere Intelligenz.

Im Allgemeinen sind weibliche Personen eher zur freiwilligen Untersuchungsteilnahme bereit als männliche Personen.

Siehe auch Rezeption einer aktuelleren Studie: <https://twitter.com/rubenarslan/status/1700049223889101230>:
Freiwillige gehen eher wählen und antworten eher "Ja" auf Fragen

Sampling Plan (Rekrutierungsplanung)

Vorab-existierende Daten

Wenn für eine Studie auf einem bereits bestehenden Datensatz Berechnungen durchgeführt werden, muss dies auch durchdacht und dokumentiert werden.

In Hinblick auf eine Präregistrierung wird i.d.R. auch der Zugriff konkret dokumentiert:

- **Registrierung vor der Sammlung der Daten:** Daten noch nicht erhoben, erstellt oder realisiert worden
- **Registrierung vor jeglicher menschlicher Beobachtung der Daten:** Daten vorhanden, wurden aber noch von niemandem quantifiziert, erstellt, beobachtet oder gemeldet
- **Registrierung vor dem Zugriff auf die Daten:** Daten vorhanden, aber Sie oder Ihre Mitarbeiter haben noch keinen Zugriff darauf
- **Registrierung vor der Analyse der Daten:** Daten vorhanden und Sie haben darauf zugegriffen, aber es wurde noch keine Analyse im Zusammenhang mit dem Forschungsplan durchgeführt
- **Registrierung nach der Analyse der Daten:** Sie haben auf einen Teil der für den Forschungsplan relevanten Daten zugegriffen und diese analysiert

siehe "Preregistration of secondary data analysis: A template and tutorial":

<https://research.tilburguniversity.edu/en/publications/preregistration-of-secondary-data-analysis-a-template-and-tutorial>

Sampling Plan (Rekrutierungsplanung)

Bewertungsschema

Nachvollziehbare Powerberechnung mit allen notwendigen Angaben	0	Keine Poweranalyse, bzw. unvollständige Angaben der Annahmen
	1	<p>Poweranalyse ist plausibel, die angenommene Effektstärke begründet (oder andere gute Begründung). Alle notwendigen Angaben sind gemacht (α, β bzw. Power, erwartete Effektstärke, welches Modell wurde genommen).</p> <p>Fazit: Die Poweranalyse sollte mit den Angaben vollständig reproduzierbar sein.</p> <p>Oder: Es wurde anderweitig begründet inwiefern diese Stichprobengröße zu einer informativen Studie führt.</p>

Sampling Plan (Rekrutierungsplanung)

Präregistrierung

M3 Sample size, power and precision
(1) Relevant sample sizes: e.g., single groups, multiple groups, and sample sizes (or sample ranges) found at each level of multilevel data. (2) Provide power analysis (e.g., power curves) for fixed-N designs. For sequential designs, indicate your 'stopping rule' such as the points at which you intend to be viewing your data <u>and in any way</u> analyzing them (e.g., t-tests and correlations, but even descriptively such as with histograms).
<p>Hier die Sensitivitätspoweranalyse berichten.</p>

M4 Participant recruitment, selection, and compensation
Indicate (a) methods of recruitment (e.g., subject pool advertisement, community events, crowdsourcing platforms, snowball sampling); (b) selection and inclusion/exclusion criteria (e.g., age, visual acuity, language facility); (c) details of any stratification sampling used; (d) planned participant characteristics (gender, race/ethnicity, sexual orientation and gender identity, SES, education level, age, disability or health status, geographic location); (e) compensation amount and method (e.g., same payment to all, pay based on performance, lottery).
<p><i>Participants will be recruited from the course "scientific working" in the winter semester 2023/24 at Charlotte Fresenius University Munich. All attendees and the course instructor will be invited and encouraged to participate (except for the study authors). No special selection or stratification will be applied. As compensation quantitative feedback on one's own results may be obtained.</i></p>

M5 How will participant drop-out be handled?
Indicate any special treatment for participants who drop out (e.g., there is follow-up in a manner different from the main sample, last value carried forward) or whether participants are replaced.
<p><i>No special treatment for participants who drop out.</i></p>

Praxis: Studiendesign & Analyseplan entwickeln

Schritt 1: Wie sieht Ihr Analyseplan für Ihre Hypothese(n) aus?

- **Welches Testverfahren nutzen Sie?** Welche Rolle spielt die Überprüfung von Annahmen für dieses Testverfahren (*hier z.B. aufnehmen, was Sie in Quantitative Methoden 1/2 für dieses Verfahren gelernt haben*)?
- Was ist/sind Ihre **UV(s)/Prädiktor(en)**?
 - Auf welche dieser Variablen bezieht sich die Hypothese?
 - Sind es spezielle Prädiktoren (z.B. Moderatoren? Kontrollvariablen?)
 - Falls zutreffend: Was ist die Begründung für die Aufnahme von Kontrollvariablen?
- Was ist Ihre **AV** bzw. Ihr **Outcome**?
- **Auf Basis welches Ergebnisses würden Sie die H0 verwerfen und die H1 annehmen** (z.B. welcher Posthoc-Test, welcher Prädiktor muss mit welchem Vorzeichen auf welchem Alpha-Niveau signifikant sein)?
- **Abgleich mit der Hypothese:** Prüft das statistische Verfahren die aufgestellte Hypothese? Muss ggf. noch etwas präzisiert werden (z.B. Form der Interaktion, ggf. (auch händische) Visualisierung der Interaktion)?
- Welche **Effekstärke** ist für Ihre geplante Analyse interessant?
- Ist ein **Manipulationscheck** geplant und wenn ja, wie wird er ausgewertet?
- Müssen Sie für **multiples Testen** korrigieren, z.B. auch weil Sie die gleiche Hypothese für verschiedene Outcomes überprüfen?

Praxis: Studiendesign & Analyseplan entwickeln

Schritt 2: Wie sieht Ihre Power bei dem vorgegebenen Samplingplan aus?

- Welche Effektstärke erwarten Sie?
- Welches Alpha-Niveau nutzen Sie? Testen Sie einseitig oder zweiseitig (gerichtete vs. ungerichtete Hypothese)?

Schritt 3: Erweitern Sie das Präregistrierungstemplate auf studynet

- Studiendesign: M6, M10, M11, M14
- Samplingplan: M3-M5 (Für M4 und M5 siehe Vorschlag von Folien)
- Analyseplan: M7-M8 (siehe Folien), M12, AP1-AP2 (siehe Folien), AP6-AP8 (AP8 optional, nur wenn Sie was exploratives unter I4 angegeben haben)

Beschreibung der konkreten Operationalisierung von Konstrukten z.B. über Fragebögen oder in Tests (Aspekt c von M12, M13, AP3) erst nächste Woche!