

Methodology: The Science before Statistics

Jede statistische Modellierung gewinnt an Aussagekraft, je umfassender sie die inhaltliche Fragestellung abzubilden im Stande ist. Um aus der riesigen Fülle an Optionen geeignet und zielgerichtet auswählen zu können sind die folgenden Unterscheidungen hilfreich.

Erkenntnisinteressen

Ganz grundlegend kann a priori das Erkenntnisinteresse von Studien in die folgenden vier Kategorien unterschieden werden:

Erkenntnisinteressen nach (Döring und Bortz 2016).			
Deskriptiv	Explorativ	Explanativ	Prädiktiv
populationsbeschreibend	hypothesengenerierend	hypothesenprüfend	Datenpunkte vorhersagend oder imputierend
Bei welchem Anteil 15-Jähriger in Deutschland handelt es sich um funktionale Analphabet:innen?	Was sind potentielle Ursachen für genderbezogene Disparitäten im Analphabetismus?	Sind 15-jährige Jungen häufiger Analphabeten als 15-jährige Mädchen?	Mit welchen Variablen können Schüler:innen at risk erfolgreich identifiziert werden?

Gütekriterien wiss. Erkenntnis nach Campbell (1957)

Für ein erfolgreiches Studiendesign und die anschließende statistische Analyse ist es sehr wertvoll sich vorab über Schwerpunkte besonders gewünschter Aspekte wissenschaftlicher Güte Gedanken zu machen. Insbesondere über die Unterkriterien **Methodischer Strenge**

- **Konstruktvalidität** (Inwiefern ist die Interpretation der Messwerte angemessen?)
- **Interne Validität** (Inwiefern sind Assoziationen von unabhängiger [beeinflussender] und abhängiger [beeinflusster] Variabler als kausale Effekte interpretierbar?)
- **Externe Validität** (Inwiefern können die Schlussfolgerungen der Studie verallgemeinert werden?)
- **Statistische Validität** (Wie robust und angemessen sind die verwendeten statistischen Verfahren?)

Deskriptiv- und Inferenzstatistik

Deskriptive Statistik beschreibt vorliegende Daten (z.B. Effektstärken), während **Inferenzstatistik** Aussagen über den die Daten generierenden Mechanismus trifft. Beide können »eher einfach« oder »hoch komplex« sein und oftmals sethen sie in einem synergetischen Verhältnis (siehe [Abbildung 1](#)).



Abbildung 1: Verhältnis von deskriptiver und Inferenzstatistik"

Bayesianisches und Frequentistisches Schätzen und Testen

Eine sehr heuristische Klassifikation inferenzstatistischer Verfahren stellt die Unterscheidung von statistischer Schätzung und Testung:

(Inferenzstatistische) <i>Schätzungen</i> (estimation with quantified uncertainty) treffen anhand von Stichproben Aussagen über Parameter der Grundgesamtheit (Population) aus der die Stichprobe gezogen wurde. ¹	¹ Bspw.: Mit 96%er Wahrscheinlichkeit liegt die Analphabetismuszinzidenz von 15-Jährigen in Deutschland zwischen .08 und .12
(Inferenzstatistische) <i>Hypothesentests</i> bewerten anhand von Stichprobendaten die Gültigkeit von Hypothesen in der Grundgesamtheit (Population) aus der die Stichprobe gezogen wurde. ²	² Bspw.: Nimmt man an, dass sich die Analphabetismuszinzidenz von 15-Jährigen in Deutschland zwischen 2021 und 2025 nicht geändert hat beträgt die Wahrscheinlichkeit der vorliegenden Daten $p = .032$

Diese beiden Verfahren können sowohl im Rahmen der **frequentistischen Statistik** als auch der **bayesianischen Statistik** angewendet werden. Die folgende Tabelle gibt einen Überblick über die wichtigsten Werkzeuge:

	Frequentistische Statistik	Bayesianische Statistik
Parameterschätzung	Konfidenzintervalle	Posterior Distributions
Hypothesentest	p-Werte	Bayes Faktoren & ROPE Procedure

Hypothesenarten

Bayesianische wie frequentistischen Hypothesentests können unterschiedliche Arten von Hypothesen zugrunde gelegt werden:

- **Punkthypothesen** setzen Parameter gleich einer reellen Zahl; etwa $H_0: \delta = 0$
- **Äquivalenzhypothesen** nehmen Parameter in einem reellen Intervall an; etwa $H_0: \delta \notin [-.3, .3]$
- **Informative Hypothesen** nehmen eine Ordnungsrelation mehrerer Parameter an; etwa $\mu_{Baseline} < \mu_{Imaginary Pill} < \mu_{Blinded Placebo}$ (Buerger u. a. 2023)

Die Art der (falsifizierten) Hypothese entscheidet wesentlich stärker über den Informationsgehalt eines Hypothesentests als die Entscheidung für das frequentistische oder bayesianische Paradigma (Hojtink 2012).

Dies ist am leichtest anhand der Nullhypothese nachvollziehbar. Wird etwa die Nullhypothese $H_0: \delta = 0$ verworfen, wird entsprechend die Alternativhypothese $H_A: \delta \neq 0$ angenommen. Diese enthält aber quasi keine Information, da sie nur mit einer einzigen Beobachtung ($d = 0.000000 \dots$) verworfen werden kann und im kritischen Rationalismus gilt, dass eine Aussage umso mehr Information enthält, umso leichter sie verworfen werden kann (Döring und Bortz 2016).

Äquivalenzhypothesen können sowohl frequentistisch (z.B. TOAST-Prozedur in R und JASP, Lakens 2017) wie bayesianisch (z.B. ROPE-Ansatz Kruschke 2015) getestet werden. Für das Testen informativer Hypothesen liegen bayesianische Methoden in (u.a.) JASP und R vor (z.B. {bain}, Gu u. a. 2019) sowie in frequentistischen R-Paketen restriktor (Vanbrabant 2020) und ic.infer (Grömping 2010).

Grundlagen der Regressionsanalyse

Regressionsanalysen sind ein sehr mächtiges Werkzeug um Zusammenhänge oder Unterschiede zwischen/in Variablen zu modellieren. Innerhalb der Regressionsanalyse kann

Literatur

Buerger, Sarah, Dilan Sezer, Niels Bagge, Irving Kirsch, Cosima Locher, Claudia Carvalho, und Jens Gaab. 2023. „Imaginary Pills and Open-Label Placebos Can Reduce Test Anxiety by Means of Placebo Mechanisms Scientific Reports”. *Scientific Reports* 13 (1): 2624. <https://doi.org/10.1038/s41598-023-29624-7>.

Campbell, Donald T. 1957. „Factors Relevant to the Validity of Experiments in Social Settings”. *Psychological Bulletin* 54 (4): 297–312.

<https://doi.org/10.1037/h0040950>.

Döring, Nicola, und Jürgen Bortz. 2016. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5., vollst. Berlin, Heidelberg: Springer.

Grömping, Ulrike. 2010. „Inference with Linear Equality and Inequality Constraints Using R: The Package lc.infer”. *Journal of Statistical Software* 33 (Februar): 1–31. <https://doi.org/10.18637/jss.v033.i10>.

Gu, Xin, Herbert Hoijtink, Joris Mulder, und Yves Rosseel. 2019. „Bain: A Program for Bayesian Testing of Order Constrained Hypotheses in Structural Equation Models”. *Journal of Statistical Computation and Simulation* 89 (8): 1526–53. <https://doi.org/10.1080/00949655.2019.1590574>.

Hojtink, Herbert. 2012. *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences Series. Boca Raton: CRC.

Kruschke, John K. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. 2nd Aufl. Academic Press.

Lakens, Daniël. 2017. „Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses”. *Social Psychological and Personality Science* 8 (4): 355–62. <https://doi.org/10.1177/1948550617697177>.

Vanbrabant, L. 2020. *Restriktor: Constrained Statistical Inference*. <https://cran.r-project.org/package=restriktor>.